

Univerzita Karlova

Filozofická fakulta

Ústav informačních studií a knihovnictví

Diplomová práce

Bc. Dominik Bláha

Analýza pro vytvoření institucionálního repozitáře na České zemědělské univerzitě v Praze

Analysis for the building of the institutional repository for the Czech
University of Life Science in Prague

Praha 2017

Vedoucí práce: PhDr. Hana Landová, Ph.D.

Poděkování

Rád bych poděkoval PhDr. Haně Landové, Ph.D. za odborné konzultace, cenné rady a trpělivost. Dále bych chtěl poděkovat své rodině za podporu, svým přátelům za péči o mé duševní zdraví a své přítelkyni za motivaci a důvěru v mé schopnosti.

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

V Praze, dne 26. července 2017

.....

Jméno a příjmení

Klíčová slova

Česká zemědělská univerzita v Praze, institucionální repozitář, důvěryhodný institucionální repozitář, OpenDOAR, ROAR, PLATTER, Data Seal of Approval, Nestor Seal of Trustworthy Digital Archives, ISO 16363, ISO 14721

Keywords

Czech University of Life Sciences, institutional repository, trustworthy institutional repository, OpenDOAR, ROAR, PLATTER, Data Seal of Approval, Nestor Seal of Trustworthy Digital Archives, ISO 16363, ISO 14721

Abstrakt

Diplomová práce má za cíl za pomoci analýzy současné situace institucionálních repozitářů v České republice navrhnout postup pro implementaci institucionálního repozitáře na České zemědělské univerzitě v Praze. Výchozím bodem pro výběr zkoumaných repozitářů jsou registry OpenDOAR a ROAR. V první části práce jsou popsány použité softwary pro implementaci repozitářů v Česku. Následuje popis nástrojů, standardů a certifikací relevantních pro důvěryhodné institucionální repozitáře jako jsou DRAMBORA a PLATTER, standardy ISO 16363 a ISO 14721 a také tzv. nižší certifikace Data Seal of Approval a Nestor Seal of Trustworthy Digital Archives. V další části jsou přiblíženy politiky a provoz zkoumaných institucionálních repozitářů. Poslední část práce je věnována popsání současného stavu na České zemědělské univerzitě, provedenému kvantitativnímu průzkumu mezi vědeckými pracovníky na této instituci a samotnému návrhu implementace institucionálního repozitáře za využití nástroje PLATTER. Závěr práce obsahuje shrnutí problematiky institucionálních repozitářů.

Abstract

The goal of the diploma thesis is to put forward a procedure for implementation of an institutional repository at the Czech University of Life Sciences in Prague with the aid of an analysis on the current situation of institutional repositories in Czechia. The analysed repositories are selected using OpenDOAR and ROAR registries. The first part of the thesis describes the software used to implement the institutional repositories in Czechia. In the next part tools, standards and certifications relevant for trustworthy institutional repositories such as DRAMBORA and PLATTER, standards ISO 16363 and ISO 14721 and the so called lesser certifications Data Seal of Approval and Nestor Seal of Trustworthy Digital Archives are described. Following part focuses on policies and operation of those analysed repositories. The last part of the thesis deals with a description of current practice on the Czech University of Life Sciences, the conducted quantitative analysis among the researchers of the university and the aforementioned procedure for implementation of an institutional repository using the tool PLATTER. The conclusion summarizes the issues of institutional repositories in Czechia.

Obsah

1	Úvod	10
1.1	Historie institucionálního repozitáře	11
1.1.1	ArXiv.org	11
1.1.2	OAIS, DSpace a SPARC	11
1.2	Současnost institucionálního repozitáře	12
1.2.1	Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH) ...	12
1.2.2	Důvěryhodný repozitář	14
1.2.3	Linked data a linked open data	15
2	Software pro institucionální repozitáře v ČR	18
2.1	Registry otevřených repozitářů	18
2.2	Software pro institucionální repozitáře v České republice podle OpenDOAR	19
2.3	Software pro institucionální repozitáře v České republice podle ROAR	19
2.4	Představení softwaru využíteho pro repozitáře v České republice	20
2.4.1	DSpace	20
2.4.2	Invenio	23
2.4.3	ARL	24
2.4.4	Digitool	25
2.4.5	EPrints	25
2.4.6	Informační systém Masarykovy univerzity	26
3	Vybrané standardy a certifikace pro institucionální repozitáře	27
3.1	Historický vývoj standardů a certifikací	27
3.2	Důvěryhodný repozitář	29
3.3	Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)	30
3.4	Catalogue of Criteria for Trusted Digital Repositories	31
3.5	Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) ...	32
3.6	Planning Tool for Trusted Electronic Repositories (PLATTER)	33

3.7	Data Seal of Approval	35
3.8	Nestor Seal for Trustworthy Digital Archives	37
3.9	Audit and Certification of Trustworthy Digital Repositories	39
4	Politiky institucionálních repozitářů v ČR.....	41
4.1	Repozitář LINDAT/CLARIN	43
4.2	Digitální repozitář Univerzity Karlovy	43
4.3	Digitální univerzitní repozitář Univerzity Karlovy	44
4.4	Česká matematická digitální knihovna	44
4.5	Digitální knihovna Vysokého učení technického	45
4.6	Digitální knihovna Českého vysokého učení technického.....	46
4.7	Digitální knihovna Filozofické fakulty Masarykovy univerzity	46
4.8	Univerzitní repozitář Masarykovy univerzity	47
4.9	Digitální knihovna Univerzity Pardubice.....	47
4.10	Digitální repozitář Vysoké školy báňské - Technické univerzity Ostrava	48
4.11	Repozitář publikační činnosti Univerzity Tomáše Bati ve Zlíně	49
4.12	Národní úložiště šedé literatury	50
4.13	Studijní knihovna pro výzkum vývoje NTK	51
4.14	Repozitář Akademie věd České republiky	51
4.15	Digitální knihovna Západočeské univerzity v Plzni.....	52
5	Současný stav na České zemědělské univerzitě v Praze.....	53
5.1	Institucionální repozitář na České zemědělské univerzitě	53
5.2	Dotazníkové šetření mezi zaměstnanci a doktorandy ČZU	54
5.3	Výběr otázek	55
5.4	Metodologie	57
5.4.1	Získání dat a stanovení hypotéz.....	58
5.4.2	Úprava dat.....	59
5.4.3	Propagace dotazníku a výběr vzorku	59

6	Návrh postupu pro tvorbu institucionálního repozitáře na ČZU	61
6.1	Výsledky průzkumu na ČZU	61
6.2	Stanovení klíčových skupin pro fungování repozitáře.....	63
6.3	Stanovení základních parametrů repozitáře	63
6.3.1	Účel a funkce repozitáře	64
6.3.2	Velikost repozitáře.....	64
6.3.3	Provoz.....	66
6.3.4	Technická řešení a možnosti implementace	67
6.4	Plány strategických cílů podle PLATTERu.....	68
7	Závěr.....	70
8	Seznam použité literatury	72
	Seznam ilustrací a tabulek	76
	Ilustrace.....	76
	Tabulky	76
	Příloha 1 – Otázky dotazníku pro vědecké pracovníky na ČZU	I
	Příloha 2 – Výpočet chí-kvadrátu pro vyloučení H_0	IX

Předmluva

Teoretická část této práce si klade za cíl zmapování existujících institucionálních repozitářů v České republice za využití dat z registrů OpenDOAR a ROAR. Na základě těchto dat bude poté získáno povědomí o využívaném software pro institucionální repozitáře v ČR, které bude následováno shrnutím základních informací o zjištěných softwarech a nastavením politik na konkrétních nalezených institucích. V další části budou rozebrány některé standardy a certifikace pro institucionální repozitáře s důrazem na ty, které jsou přeloženy do českého jazyka.

Praktická část této práce a její přínos je návrh postupu pro tvorbu institucionálního repozitáře na České zemědělské univerzitě v Praze (dále ČZU). Tento návrh bude vycházet z analýzy současných politik a procesů užívaných na jiných univerzitách v České republice provedené v první části. Zároveň bude reflektovat potřeby a přání členů akademické obce ČZU díky provedenému dotazníkovému průzkumu mezi doktorandy a vědci na ČZU.

Práce požadovaného rozsahu¹ si neklade za cíl detailní analýzu institucionálních repozitářů v Česku, ani popis jejich politik nad rámec zmapování současného stavu. Autor nekontaktoval správce daných repozitářů pro korekci zjištěných politik a nastavení, a zjišťoval informace výhradně z dostupných veřejných zdrojů – webových stránek repozitářů a spravujících institucí. Autor je toho názoru, že jednou z vlastností důvěryhodného institucionálního repozitáře je také přehledná prezentace jeho politik a transparentní činnost.

Bibliografické odkazy a citace v této práci se řídí normou ISO 690 a pro citování a odkazování byl zvolen Harvardský systém. Doplnující odkazy a poznámky k textu jsou uvedeny jako poznámky pod čarou na příslušných stránkách práce.

¹ 129 012 znaků s mezerami, v přepočtu 71 normostran

1 Úvod

Institucionální repozitář je odborný knihovnický termín a dle České terminologické databáze knihovnictví a informační vědy (TDKIV) má následující výklad: (Cubr a Havlová, 2003)

„Digitální repozitář obsahující digitální dokumenty, které jsou produktem výzkumné, vědecké, vývojové a další tvůrčí činnosti konkrétní instituce nebo konkrétního konsorcia.“

Hned první slovo definice je velmi určující a informačně nabitě. V dnešní digitální době totiž vzniká převážná většina vědeckých výstupů digitálně a je tedy už nemyslitelné, zabývat se klasickými repozitáři fyzických dokumentů.

Tato definice ovšem nepokrývá několik klíčových aspektů. První je specifický pro vzdělávací instituce a to digitální dokumenty vznikající jako finální produkt vzdělávací činnosti – kvalifikační práce. Byla by přitom škoda v institucionálních repozitářích tento typ dokumentů opomenout, protože sám o sobě má jistou informační hodnotu. Schopfel et al. (2014) zkoumá přínos dodatečných materiálů nashromážděných pro potřeby kvalifikačních prací, zejména u disertačních prací a vyvozuje, že tyto práce jsou důležitou součástí institucionálních repozitářů. Kvalifikační práce také mohou udávat trend, kterým se ubírá výzkum na dané instituci. Na druhou stranu jazyk, jakým jsou tyto práce psány, a jejich popisnost, může být mnohem přístupnější neodborné veřejnosti a pomoci propagaci vědy.

Další aspekt, který tato definice nepokrývá je problematika otevřeného přístupu. V dnešní době se totiž stále více užívá termínu institucionální repozitář ve významu otevřený repozitář a spolu s termínem otevřená věda. Repozitáře totiž umožňují celkem jednoduše rozšiřovat výsledky vědeckého výzkumu a to nezávisle na zavedených vydavatelích a jejich většinou netransparentních procesech. Díky převládajícímu chápání termínu institucionální repozitář ve vztahu k otevřenému přístupu se bude tato práce zabývat institucionálními repozitáři ve smyslu otevřený institucionální repozitář.

Posledním aspektem zcela opomenutým touto definicí je problematika dlouhodobé udržitelnosti obsahu institucionálního repozitáře. Kvalifikační práce mají prozatím vždy povinný výtisk, který putuje do institucionálního archivu či knihovny, nicméně například u odborných článků tomu tak není a nelze se spoléhat na vydavatele s výlučnými právy k obsahu nebo autora, že se o dlouhodobé uchování dokumentu postarají.

1.1 Historie institucionálního repozitáře

Díky rozmachu open source a možnostem jednoduše v tomto režimu vytvářet software a možná i díky specifickému prostředí vědeckých a vzdělávacích institucí je dnes na trhu mnoho různých webových aplikací pro realizaci institucionálního repozitáře. Díky komplexnosti dostupných nástrojů je dnes prakticky možné, že instituce je tvůrcem, vydavatelem i distributorem výsledků vědeckého bádání a to včetně recenzování i archivace. Nicméně nebylo tomu tak vždy.

1.1.1 ArXiv.org

V roce 1991 byl spuštěn první institucionální repozitář známý jako arXiv.org². Tento repozitář původně sloužil jen pro potřeby fyziků bádajících v oblasti částicové fyziky (high energy physics), nicméně zanedlouho byl rozšířen i pro obory matematiky, počítačových věd, statistiky a dalších. To, co jej dělalo a dělá jedinečným je způsob distribuce a vlastně už samotný proces publikace vědeckých výsledků. (Cornell University Library, 2017b)

Během publikování může vzniknout více verzí dané publikace. Kdykoli totiž autor nahraje novou verzi svého rukopisu, není původní verze ztracena, ale zůstává uchována v repozitáři. Publikace jsou volně dostupné online komukoli bez přihlášení a jakéhokoli příspěvku. V popisu mise repozitáře jeho správci jasně stanovují závazek k zajištění trvalého přístupu ke všem zveřejněným publikacím (Cornell University Library, 2017a).

1.1.2 OAIS, DSpace a SPARC

Rok 2002 byl zlomovým pro budoucí rozvoj institucionálních repozitářů. Hned v lednu 2002 byla vydána klíčová publikace pro sjednocení způsobu, jakým budou repozitáře fungovat – Reference Model for an Open Archival Information System (OAIS) (The Consultative Committee for Space Data Systems, 2012). Ve své první verzi Blue Book³, se jednalo více méně o standardizaci názvosloví a infrastruktury a stanovení požadavků na OAIS. Později v roce 2012 vyšlo nové vydání Magenta Book, které již obsahuje mimo standardů také doporučení a postupy, jak OAIS systém implementovat. Protože toto nové vydání nahradilo původní dokument, cituji pouze aktuální verzi. Tento dokument byl později publikován jako standard ISO 14721⁴. (ISO 14721:2012, 2012)

² arXiv.org - <https://arxiv.org/>

³ Organizace CCSDS dělí své publikace do tzv. books různých barev – Blue Books jsou doporučené standardy, Magenta Books rozvíjí standardy o doporučenou implementaci standardů

⁴ ISO 14721:2012 Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model: <https://www.iso.org/standard/57284.html>

V roce 2002 byla také vytvořena první veřejná verze softwaru pro provozování institucionálního repozitáře s otevřeným zdrojovým kódem – DSpace. Kolem tohoto nástroje se vytvořila silná komunita a, i když není financován a vyvíjen konkrétní komerční organizací, dokázal se prosadit jako jedno z nejpoužívanějších řešení pro provoz institucionálního repozitáře. Bližším informacím o softwarech pro provoz institucionálních repozitářů bude věnována kapitola 2 Software pro institucionální repozitáře v ČR.

Ještě stále v roce 2002 vydala organizace SPARC⁵ publikaci SPARC Institutional Repository Checklist and Resource Guide (Crow, 2002). Tato organizace je již od začátku jedním z hlavních podporovatelů nezávislého publikování v podobě institucionálních repozitářů. Zmíněná publikace se věnuje problémům a otázkám, které je potřeba vyřešit, pokud se instituce chystá implementovat institucionální repozitář a to zejména:

- Organizační a administrativní problémy a nastavení politik
- Politiky ohledně obsahu a politiky zpřístupňování a udržování
- Zapojení tvůrců obsahu a jejich aktivní oslovení
- Technické možnosti a otázky infrastruktury

Publikace staví na zkušenostech institucí, které již repozitář implementovaly, nicméně nenabízí konkrétní řešení systémem krok za krokem, ale spíše perspektivy a možné komplikace.

Díky výše zmíněným událostem a dokumentům dostávají instituce po celém světě do rukou jak prostředek, tak i návod k tomu, jak vlastní institucionální repozitář implementovat.

1.2 Současnost institucionálního repozitáře

Zatímco se terminologie a obecná architektura repozitářů v průběhu dalších let ustalovala, vyvstaly další otázky v souvislosti s tímto fenoménem. Těmito otázkami jsou centrální vyhledávání mezi repozitáři a agregace metadat, dlouhodobá udržitelnost repozitářů (v řádech desítek let) a rozšíření záběru i na data vznikající během výzkumu.

1.2.1 Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH)

Už v začátcích repozitářů, v 90. letech 20. století, se hledaly cesty, jak by mezi sebou mohly repozitáře sdílet metadata a umožnit tak vyhledávání napříč různými instalacemi. Tak, jak se rozrůstal počet různých řešení, tato potřeba ještě narůstala, protože různé typy repozitářů měly různě navržená rozhraní stroj-člověk. Tato situace znesnadňovala řešerši ve více repozitářích,

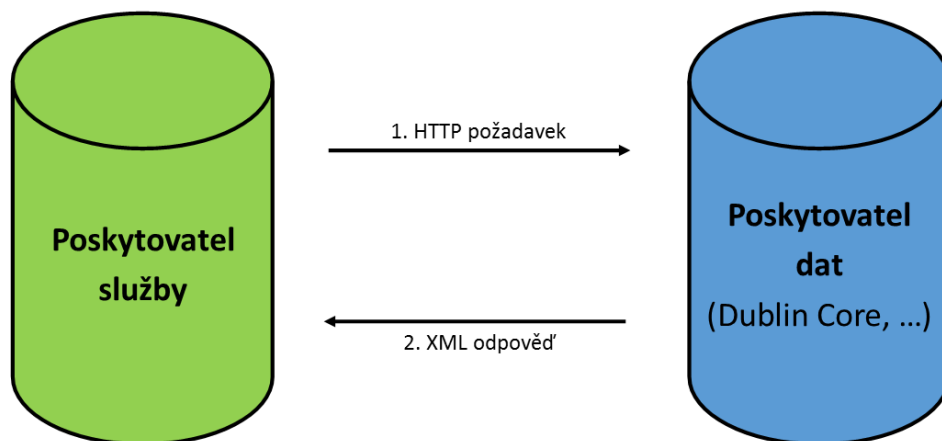
⁵ The Scholarly Publishing & Academic Resources Coalition: <https://sparcopen.org/>; případně evropská pobočka organizace SPARC Europe: <http://sparceurope.org/>

protože uživatel se musel učit pracovat s různými rozhraními. Stejně tak rozhraní pro komunikaci stroj-stroj umožňovala maximálně komunikovat mezi repozitáři stejného typu. Proto bylo v roce 1999 svoláno tzv. Shromáždění v Santa Fe⁶, které mělo začít debatu o tom, jakým směrem se vydat při implementaci interoperability mezi různými typy repozitářů (Van de Sompel a Lagoze, 2000). Výsledkem tohoto shromáždění bylo vytvoření organizace Open Archives Initiative (OAI), pod jejíž záštitou poté v roce 2001 vznikla první verze protokolu pro výměnu metadat mezi repozitáři – OAI-PMH v1.0. Tato technologie používá pro výměnu dat protokol XML, a protože v témže roce došlo k vydání nové verze XML protokolu, vyšla také nová verze OAI-PMH v1.1, která tuto změnu reflektovala. V roce 2002 vyšla verze 2.0, která sice neumožňovala zpětnou kompatibilitu s verzí 1.1, nicméně zaručovala kompatibilitu s dalšími verzemi. Poslední úprava této verze protokolu proběhla 8. ledna 2015⁷.

Protokol OAI-PMH definuje dva typy objektu v ekosystému výměny metadat. Poskytovatele dat (Data Provider), který je informačním zdrojem a implementuje OAI-PMH za účelem poskytnutí metadat o své sbírce. Dále poskytovatel služby (Service Provider), který sklízí (harvesting) metadata přes OAI-PMH a staví na nich nějakou službu s přidanou hodnotou. Na Ilustraci 1 je znázorněno schéma komunikace protokolem OAI-PMH. Protokol vyžaduje, aby poskytovatel dat umožňoval formátování metadat ve formátu Dublin Core. Ostatní formáty jsou dobrovolné. (Lagoze et al., 2002) Přínos tohoto protokolu je jak v možnosti zpřístupnit metadata z různých systémů přes jedno rozhraní, tak ve stahování pouze inkrementálních změn, což vede k efektivnější komunikaci mezi poskytovatelem služby a poskytovatelem dat.

⁶ The Santa Fe Convention for the Open Archives Initiative: http://www.openarchives.org/sfc/sfc_entry.htm

⁷ <http://www.openarchives.org/OAI/openarchivesprotocol.html>



Ilustrace 1: Schéma protokolu OAI-PMH

1.2.2 Důvěryhodný repozitář

Zatímco v počátcích repozitářů se vedla debata spíše o technickém řešení, později vyvstal problém s udržitelností již existujících repozitářů. Tedy, aby repozitář najednou nepřestal existovat se všemi daty v něm, ale mohl zaručit provoz na několik (desítek) let dopředu. TDKIV definuje důvěryhodný digitální repozitář jako (Cubr a Havlová, 2006):

„Digitální repozitář, který jasně definované skupině uživatelů poskytuje spolehlivý a dlouhodobý přístup k organizovaným digitálním sbírkám v dlouhodobém horizontu. Důvěryhodnost digitálního repozitáře musí být v současnosti možné objektivně testovat a vyhodnotit na základě předem stanovených kritérií, mezi která patří například schopnost zajistit kontinuitu financování veškerého provozu. Testování a vyhodnocení může vykonávat externí subjekt.“

Vytvoření důvěryhodného repozitáře neznámá jen zajištění financování. Rosenthal et al. (2005) ve své rané práci ještě nepoužívá termín důvěryhodný digitální repozitář, nicméně zaměřuje se na aspekty dlouhodobé archivace obecně a stanovuje hrozby, kterým musí umět repozitáře čelit. Mimo to navrhuje také strategie, jakými by se pokrytí jím zmíněných hrozeb dalo provést. První je replikace, tedy ukládání několika kopií té samé informace. Druhou je migrace, což znamená uložení té samé informace na různé typy médií (magnetické pásky, optické disky a další). Třetí navrhovanou strategií je transparentnost. Zejména ve smyslu zpřístupnění informací o tom, jak se nakládá s nahrávaným obsahem, licencování dat, dokumentace, apod. Dále uvádí diverzitu, auditování, snížení ekonomické náročnosti příjmu,

ukládání a zpřístupňování obsahu, a nakonec i zpomalení akcí, které nemusí být provedeny ihned, aby byl dostatek času na odhalení a opravení případné chyby.

Očividně nemůže být důvěryhodnost repozitáře samodeklarovaná a jak píše citovaná definice výše, je potřeba stanovit kritéria pro její určení a kontrolu. Za tímto účelem vzniklo hned několik certifikací pro digitální repozitáře – některé komerční, jiné pro vlastní vyhodnocení správci příslušného repozitáře. Mezi nimi například Trustworthy Repositories Audit & Certification (TRAC), ISO 16363:2012, PLATTER, DRAMBORA nebo Data Seal of Approval. O těchto certifikacích pojednává kapitola 3 Vybrané standardy a certifikace pro institucionální repozitáře.

1.2.3 Linked data a linked open data

Samotný internet, který je dnes nedílnou součástí naší kultury, prochází vývojem. Další vývojový stupeň dnešního internetu je sémantický web. Původní web, tak jak jej známe a desítky let využíváme, je tvořen dokumenty, které na sebe odkazují pomocí hypertextových odkazů. Nicméně toto řešení není příliš vhodné pro strojové zpracování a většinou vyžaduje člověka pro určení kontextu daného webu. Sémantický web stojí na těchto základech, tedy je také tvořen dokumenty vystavenými na webu. Avšak zatímco hypertextové odkazy propojují jednotlivé HTML dokumenty, v sémantickém webu na sebe odkazují koncepty a objekty popsané RDF za pomoci URI odkazů. V praxi to vypadá tak, že přímo v dokumentu se nachází HTML tagy, které standardizovaným způsobem popisují kontext dokumentu. V případě kvalifikační práce uložené v institucionálním repozitáři lze třeba z těchto tagů zjistit autora, vedoucího práce, datum publikace, typ práce apod. viz Ilustrace 2. V dnešní době je jako metadatové schéma pro tyto objekty nejčastěji používán Dublin Core.

Tato metadata jsou dostupná i v klasickém webu, ale ačkoliv my je máme na očích, pro stroje jsou paradoxně skryta. Je to z toho důvodu, že v jazyce HTML, ve kterém jsou webové stránky popsány, není možné zachytit kontext obsahu. Zároveň stroje nejsou schopny rozlišit, jaká

informace se nachází v konkrétním odstavci nebo buňce tabulky, protože jim chybí možnost porozumění textu jako nám lidem.

```
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.contributor" content="Lednický, Václav" xml:lang="en" />
<meta name="DC.creator" content="Štolcová, Zuzana" xml:lang="en" />
<meta name="DCTERMS.dateAccepted" content="2006-04-20"
scheme="DCTERMS.W3CDTF" />
<meta name="DCTERMS.available" content="2006-04-20"
scheme="DCTERMS.W3CDTF" />
<meta name="DCTERMS.issued" content="2005" xml:lang="en"
scheme="DCTERMS.W3CDTF" />
<meta name="DC.identifier" content="200515851 DP" xml:lang="en" />
<meta name="DC.identifier" content="http://hdl.handle.net/10084/36775"
scheme="DCTERMS.URI" />
<meta name="DC.format" content="59 l., [10] l. příl. : il."
xml:lang="en" />
<meta name="DC.language" content="cs" xml:lang="en"
scheme="DCTERMS.RFC1766" />
<meta name="DC.title" content="Cestovní ruch na Šumpersku" xml:lang="cs"
/>
<meta name="DC.type" content="Diplomová práce" xml:lang="en" />
<meta name="DC.identifier" content="141883/15851" xml:lang="en" />
<meta name="DC.identifier" content="ÚK/Sklad diplomových prací"
xml:lang="en" />
<meta name="DC.date" content="2005" xml:lang="en"
scheme="DCTERMS.W3CDTF" />
<meta name="DC.description" content="Prezenční výpůjčka" xml:lang="en"
/>
<meta name="DC.description" content="VŠB - Technická univerzita Ostrava.
Ekonomická fakulta. Katedra (115) podnikatelství a managementu"
xml:lang="cs" />
<meta content="Štolcová, Zuzana" name="citation_author">
<meta content="2005" name="citation_date">
<meta content="http://dspace.vsb.cz/handle/10084/36775"
name="citation_abstract_html_url">
<meta content="cs" name="citation_language">
<meta content="Diplomová práce" name="citation_keywords">
<meta content="Cestovní ruch na Šumpersku" name="citation_title">
```

Ilustrace 2: Ukázka popisu dokumentu za pomoci schématu Dublin Core v hlavičce dokumentu. (Zdroj: repozitář VŠB-TUO)

Základní složkou sémantického webu představují propojená data (linked data). S vývojem konceptu propojených dat začal sám otec internetu Tim Berners-Lee. Ve svém článku (Berners-Lee, 2010) jednak popisuje linked data a také stanovuje 4 základní pravidla pro jejich úspěšné nasazení a efektivní využití:

1. Použití URI jako jméno pro objekty.
2. Použití http URI, aby si lidé mohli vyhledat tato URI jména.
3. Když si někdo vyhledá URI, poskytněte užitečné informace za využití standardů (RDF).
4. Přidejte odkazy na související URI objekty, aby byla zajištěna propojenost objektů.

Pojem linked open data označuje data, která jsou propojena pomocí URI a navíc poskytována k veřejnému užití pod nějakou otevřenou licenci. Rösslerová (2016) ve své disertační práci mapující výměnné formáty dat z pohledu knihovnictví uvádí: „Z těchto analýz je dále zřejmé, že knihovníci doporučují a předpokládají využití propojených dat, a to pro oblast publikování dat na webu, často ale i pro samotnou výměnu nebo i pro katalogizaci.”

2 Software pro institucionální repozitáře v ČR

V současnosti existují jistě desítky implementací softwaru institucionálního repozitáře. Obzvláště přihlédneme-li k faktu, že každá větší univerzita má nějaké IT oddělení a není zdaleka pravidlem, že by se vždy sáhlo po řešení, které již existuje. Podle dat z Ministerstva školství, mládeže a tělovýchovy (2015) bylo k 31. 12. 2015 dohromady v České republice 67 veřejných a soukromých vysokých škol. Tato diplomová práce si nedává za cíl podrobně zmapovat veškerý software pro repozitáře používaný v ČR, a proto je k získání přehledu o používaném softwaru pro institucionální repozitář v ČR využito dat ze dvou dnes nejrozšířenějších mezinárodních registrů otevřených repozitářů – OpenDOAR⁸ a ROAR⁹.

2.1 Registry otevřených repozitářů

Registr OpenDOAR je výsledkem spolupráce Nottinghamské univerzity ve Velké Británii a Lundské univerzity ve Švédsku a byl založen v roce 2003. Jeho vznik iniciovala potřeba mít nejen soupis repozitářů existujících po celém světě, ale také uchovávat a zpřístupňovat i další informace o těchto repozitářích jako: umístění, obsah, použitý software, politiky, tematický záběr, adresu OAI-PMH serveru atd. V počátcích měl tento registr hned několik finančních sponzorů a to organizace OSI¹⁰, Jisc¹¹, SPARC Europe¹² a RLUK¹³. Dnes je provoz a další rozvoj pokrýván pouze organizací Jisc a financováním z vlastních zdrojů. Díky množství shromažďovaných dat, dobré infrastruktuře, nástrojům a licencování uložených dat licencí Creative Commons BY-NC-SA 3.0¹⁴ lze nad obsahem tohoto registru snadno provádět výzkumy a analýzy. (University of Nottingham, 2014)

Druhým registrem je ROAR (Registry of Open Access Repositories). Na počátku, v roce 2004, byl tento registr vytvořen pouze za účelem shromáždění seznamu a informací o repozitářích vybudovaných na softwaru EPrints. Nicméně později byl rozšířen také o informace o repozitářích běžících i na jiných softwarech. Registr je provozován na platformě EPrints a

⁸ Directory of Open Access Repositories: <http://www.opendoar.org/>

⁹ Registry of Open Access Repositories: <http://roar.eprints.org/>

¹⁰ Open Society Institute do r. 2011, dnes Open Society Foundations: <https://www.opensocietyfoundations.org/>

¹¹ Dříve Joint Information Systems Committee, dnes pouze Jisc: <https://www.jisc.ac.uk/>

¹² SPARC Europe: <http://sparceurope.org/>

¹³ Research Libraries UK: <http://www.rluk.ac.uk/>

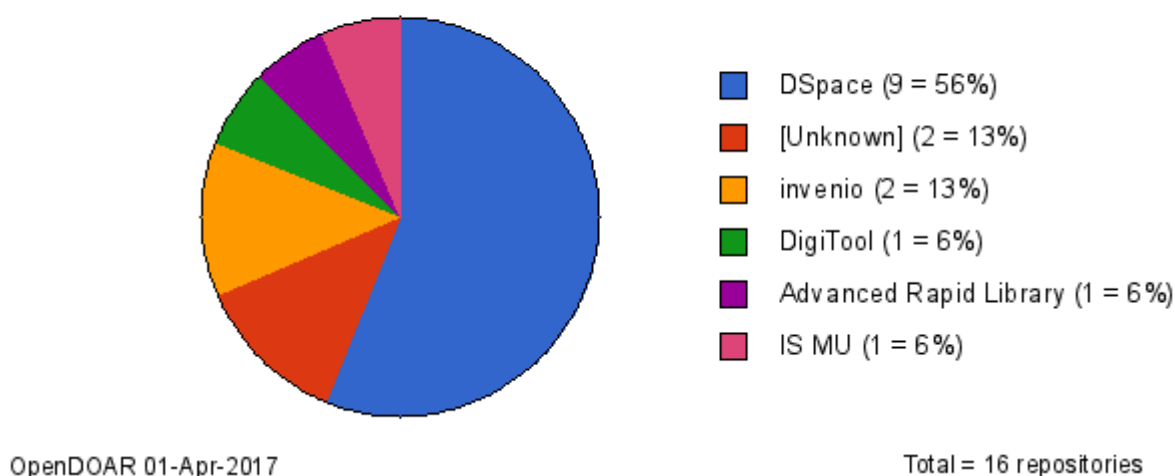
¹⁴ Bližší vysvětlení zmíněné licence na <https://creativecommons.org/licenses/by-nc-sa/3.0/>

spravuje jej Southamptonská univerzita ve Velké Británii. Financování repozitáře je zajištěno také organizací Jisc jako v případě OpenDOAR. (University of Southampton, 2004)

2.2 Software pro institucionální repozitáře v České republice podle OpenDOAR

Registr OpenDOAR poskytuje možnost vyhledávat podle kategorií jako umístění, jazyk, zaměření repozitáře, typ obsahu či použitý software. Z výsledků vyhledávání lze pak jednoduchým nástrojem vytvořit sloupcový nebo koláčový graf.

Usage of Open Access Repository Software
Czech Republic



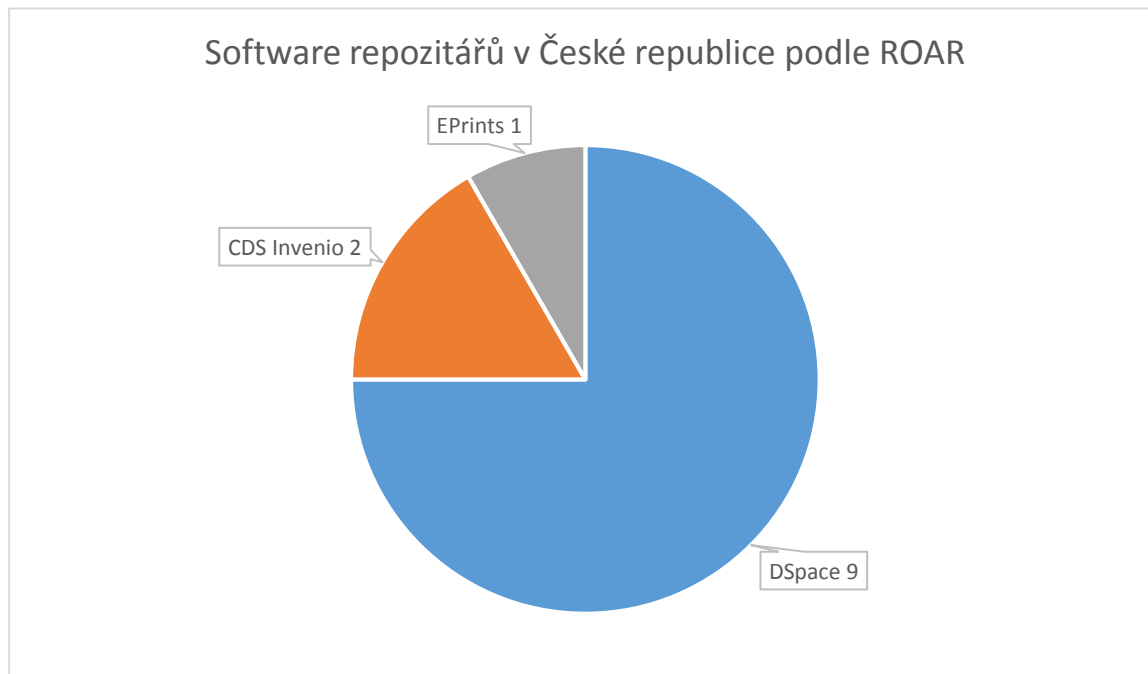
Ilustrace 3: Přehled softwaru použitého pro institucionální repozitáře v České republice podle OpenDOAR [vid. 1. 4. 2017]

Jak je patrné z Ilustrace 3, více než polovina českých repozitářů vedených v OpenDOAR je provozována na open source softwaru DSpace. Zbylé repozitáře jsou postaveny na různém software, nicméně žádný nedominuje tak výrazně jako právě DSpace. Na dvou repozitářích je open source software Invenio. Po jednom repozitáři mají komerční software DigiTool od ExLibris a Advanced Rapid Library od firmy Cosmotron. Jeden repozitář je provozován na vlastním řešení IS MU, což je informační systém Masarykovy univerzity v Brně.

2.3 Software pro institucionální repozitáře v České republice podle ROAR

ROAR umožňuje vyhledávat v názvu indexovaných repozitářů, dále popisu, typu, datu vzniku, umístění, softwaru či předmětovém zaměření. Výsledky vyhledávání lze pak pomocí nástroje převést do několika typů jednoduchých grafů či vyexportovat jako textová data. Zde bylo zvoleno exportu dat, protože nabízené grafy nebyly dostatečně informativní.

Na Ilustraci 4 je vidět, že ROAR neindexuje tolik repozitářů jako OpenDOAR. Celkem pouze 12 repozitářů oproti 16 repozitářům v OpenDOAR. Nicméně i v tomto registru jasně převládá software DSpace a v pozadí stojí Invenio a EPrints. V případě obou registrů jsou tedy více než dvě třetiny repozitářů v České republice postaveny na open source řešení.



Ilustrace 4: Graf softwaru použitého pro repozitáře v České republice podle registru ROAR [vid. 2. 4. 2017]

2.4 Představení softwaru využíteho pro repozitáře v České republice

Pro potřeby této práce budou v následující části představena různá softwarová řešení tak, jak byla zjištěna z dat z registrů OpenDOAR a ROAR. Z nalezených údajů vyplývá, že open source řešení (DSpace a Invenio) jasně dominují prostředí českých institucionálních repozitářů, avšak v krátkosti bude popsán také open source software EPrints. Pro úplnost budou představeny také nalezené nástroje komerční (ARL a DigiTool) a vlastní řešení z Masarykovy univerzity, které se v registrech objevily.

2.4.1 DSpace

Jako reakce na vzrůstající množství výstupu univerzit vznikajícího digitálně a nutnost tento výstup nejen zpřístupňovat, ale i dlouhodobě ukládat, byla v roce 2002 vytvořena první verze open source software pro institucionální repozitáře DSpace. Tento systém vznikl společnou

snahou Massachusettského technologického institutu (MIT) a Hewlett-Packard Laboratories¹⁵, vývojové a inovační divize firmy Hewlett-Packard¹⁶. (Smith et al., 2003)

V době psaní této práce je poslední verze ke stažení DSpace 6.0. Software je distribuován pod open source licencí BSD¹⁷. Z technické perspektivy je systém napsán v programovacím jazyce Java a dokáže běžet na serveru s operačním systémem Linux i Windows. Pro interakci s uživatelem i administrátory slouží webové rozhraní, které nabízí (DuraSpace Organization, 2016b):

- vyhledávání skrze Apache Solr,
- workflow pro posuzování nahrávaných dokumentů,
- možnost nahrát jakýkoliv soubor (systém umí sám rozpoznat přípony jako DOC, PDF, XLS, PPT, JPEG, MPEG či TIFF),
- metadatový popis podle schématu Qualified Dublin Core,
- rozšíření skrze pluginy,
- autentizaci a autorizaci uživatelů podle LDAP či Shibboleth,
- vícejazyčné rozhraní ve více než 20 jazycích včetně češtiny,
- sklízení dat přes OAI-PMH

Samotná aplikace běžící na pozadí umožňuje (tamtéž):

- strojový přístup k datům přes RESTful API,
- běh na databázi PostgreSQL nebo Oracle,
- ukládání dokumentů do lokálního souborového systému nebo do cloudu,
- udržování integrity nahraných souborů kontrolními součty a automatickými kontrolami

Vývoj DSpace má na starosti nezisková organizace DuraSpace ve spolupráci s dobrovolníky z celého světa. Tato spolupráce je možná díky využití vývojové platformy GitHub a podrobné a aktuální dokumentaci celého systému. O určování strategie a aktuálních priorit vývoje se stará tzv. řídicí skupina (leadership group), která je tvořena zástupci institucí a členy organizace, kteří přispěli určitou částkou. Řídicí skupina má na starosti schvalování rozpočtu a strategie, dále nominaci členů do tzv. výkonné skupiny (steering group). Výkonná skupina navrhuje

¹⁵ Dnes HP Labs: <http://www8.hp.com/us/en/hp-labs/index.html>

¹⁶ Dnes HP Inc.: <http://welcome.hp.com/country/cz/cs/cs/home.html>

¹⁷ DSpace Source Code BSD License: <http://www.dspace.org/license>

budoucí strategie, rozpočet, a má na starosti dohled nad vývojem a prezentování klíčových rozhodnutí řídicí skupině. (DuraSpace Organization, 2016a)

Mezi institucionální repozitáře v současné době využívající DSpace patří:

- Česká digitální matematická knihovna¹⁸ spravovaná Institutem matematiky na Akademii věd České republiky
- Digitální knihovna Vysokého učení technického v Brně¹⁹, která je využívána k publikaci kvalifikačních prací, publikační činnosti i některých odborných univerzitních periodik
- Digitální knihovna Českého vysokého učení technického²⁰ obsahující také kvalifikační práce, publikační činnost i některé odborné časopisy
- Digitální knihovna Filozofické fakulty Masarykovy univerzity²¹ obsahující časopisy, monografie a sborníky
- Digitální knihovna Univerzity Pardubice²², která obsahuje kvalifikační práce, publikační činnost i publikace vydané v režimu otevřeného přístupu
- Digitální repozitář Vysoké školy báňské - Technické univerzity Ostrava²³ obsahující kvalifikační práce články ze sborníku, odborná institucionální periodika a publikační činnost
- Repozitář publikační činnosti Univerzity Tomáše Bati ve Zlíně²⁴ obsahující veškerou publikační činnost autorů z univerzity, zejména hodnocené výsledky vědy a výzkumu UTB – články v časopisech, monografie a příspěvky v konferenčních sbornících
- Digitální knihovna UTB²⁵, která je vyčleněna pro kvalifikační práce instituce

¹⁸ Odkaz do repozitáře DML-CZ: <http://dml.cz/>

¹⁹ Odkaz do repozitáře VUT: <https://dspace.vutbr.cz>

²⁰ Odkaz do repozitáře ČVUT: <https://dspace.cvut.cz/>

²¹ Odkaz do repozitáře FF MU: <https://digilib.phil.muni.cz/>

²² Odkaz do repozitáře UPa: <http://dk.upce.cz/>

²³ Odkaz do repozitáře VŠB-TUO: <http://dspace.vsb.cz/>

²⁴ Odkaz do repozitáře publikační činnosti UTB: <http://publikace.k.utb.cz/>

²⁵ Odkaz do repozitáře kvalifikačních prací UTB: <http://digilib.k.utb.cz/>

- Repozitář LINDAT/CLARIN²⁶ spravovaný Ústavem formální a aplikované lingvistiky na Matematicko-fyzikální fakultě Univerzity Karlovy v Praze. Tento repozitář slouží k zpřístupňování a archivaci jakýchkoli jazykových dat a nástrojů jako korpusy, slovníky, parsery apod.
- Digitální knihovna Západočeské univerzity v Plzni²⁷ obsahující publikační činnost akademických a vědeckých pracovníků a kvalifikační práce studentů

2.4.2 Invenio

Stejně jako DSpace i Invenio bylo vytvořeno v roce 2002 jako reakce na narůstající množství digitálních dokumentů. V případě Invenia se jednalo o bibliografické záznamy dokumentů z CERNu²⁸ z oblasti částicové fyziky. O vyvinutí první verze systému se v té době postaral samotný CERN. Dnes se o vývoj systému Invenio stará mezinárodní komunita a zástupci institucí jako německé DESY²⁹, švýcarský EPFL³⁰ či americký FNAL³¹. (CERN, 2015a) Jak je vidět o vývoj se stará zejména komunita fyzikálně zaměřených institucí.

V době psaní této práce je k dispozici stabilní verze 1.2.2, která přímo vychází z prvotní verze 1.0 publikované v roce 2002. Dále je možné získat tzv. polostabilní verzi 2.1.1 a v přípravách je i verze 3.0. Software je distribuován pod licencí GNU General Public Licence (GPL). Systém je napsán v programovacím jazyce Python a je schopen provozu na serveru s unixovým operačním systémem (Linux). Pro interakci s uživateli i administrátory systému se využívá webové rozhraní nabízející:

- vyhledávání za pomoci Elasticsearch,
- správu kolekcí (na CERNu přes 1 000 000 záznamů v 700 kolekcích),
- metadatový popis podle MARC 21,
- možnost komentovat dokumenty na webu,
- rozšíření skrze moduly,
- sklizení dat přes OAI-PMH,
- procesy pro příjem záznamů dávkový import,

²⁶ Odkaz do repozitáře LINDAT/CLARIN: <https://lindat.mff.cuni.cz/repository/xmlui/?locale-attribute=cs>

²⁷ Odkaz do repozitáře ZČU: <https://dspace5.zcu.cz/>

²⁸ Conseil Européen pour la Recherche Nucléaire (European Council for Nuclear Research): <https://home.cern/>

²⁹ Deutsches Elektronen-Synchrotron (German Electron Synchrotron): <http://www.desy.de/>

³⁰ École polytechnique fédérale de Lausanne (Swiss Federal Institute of Technology in Lausanne): <https://www.epfl.ch/index.en.html>

³¹ Fermi National Accelerator Laboratory: <http://www.fnal.gov/>

- vícejazyčné rozhraní ve více než 20 jazycích včetně češtiny

Vývoj Invenia je celkem podobný vývoji DSpace. Nicméně Invenio není přímo spravováno nějakou neziskovou organizací a ani nemá odbornou komisi, která by určovala strategické cíle a další vývoj. O směřování software se totiž stará přímo CERN a nejbližší instituce zmíněné dříve. Vývoj probíhá na platformě GitHub, kde mohou programátoři jednak nahrávat svůj kód a jednak hlásit chyby či požadavky a nápady na další funkcionalitu. Invenio má pak tzv. třídící tým, který uděluje prioritu jednotlivým chybám či požadavkům a dále je třídí podle typu a nastavuje aktuální stav. (CERN, 2015b)

Invenio je v době psaní této práce nasazeno na dvou repozitářích provozovaných Národní technickou knihovnou. Jedním z nich je Institucionální digitální repozitář: Studijní knihovna pro výzkum vývoje NTK³², kde je nainstalována verze 1.0.0. (Národní technická knihovna, 2017b) Druhým repozitářem je mnohem známější Národní úložiště šedé literatury (NUŠL)³³, kde momentálně běží software Invenio ve verzi 1.1.2. V případě NUŠL je nad softwarem Invenio vystavěno přívětivější uživatelské rozhraní, které je vytvořené na míru.

2.4.3 ARL

Advanced Rapid Library je komerční software česko-slovenské společnosti Cosmotron³⁴. Je to komplexní systém určený pro správu katalogů a sbírek v knihovně v síťovém prostředí. Výrobce má certifikaci ČSN ISO 9001 a ČSN ISO 27001. Tento systém umožňuje zpracovávat nejen bibliografické informace, ale v plné míře podporuje také práci s multimediálními, obrazovými, zvukovými i fulltextovými dokumenty. (Cosmotron Bohemia s.r.o., 2014)

Tento informační systém je otevřený ve smyslu API a propojitelnosti s ostatními aplikacemi. Umožňuje strojové sklizení přes OAI-PMH, dále vyhledávání v záznamech přes protokol Z35.50 či webové služby pro komunikaci klientských aplikací s ARL založené na SRU/SRW³⁵. ARL je v současnosti využíván v repozitáři Akademie věd³⁶ pro zveřejňování jejich publikačních výstupů.

³² Odkaz do repozitáře NTK: <http://repozitar.techlib.cz/>

³³ Odkaz do repozitáře NUŠL: <http://invenio.nusl.cz/>

³⁴ Cosmotron Bohemia: <http://www.cosmotron.cz/>

³⁵ Search and Retrieve Web Service /Search and Retrieve URL Service – způsob, jak mohou klientské aplikace vyhledávat v systému za pomoci URL a XML. Nahrazuje starší Z39.50. (OCLC, 2017)

³⁶ Odkaz do Repozitáře AV ČR: <https://asep.lib.cas.cz/arl-cav/>

2.4.4 Digitool

Digitool je komerční software, který vyvinula společnost Ex Libris³⁷. Bohužel vzhledem k tomu, že tento software není již nadále vyvíjen ani nabízen, nemá smysl zacházet při jeho popisu do podrobností, i když má tento software nástupce s názvem Rosetta. Ačkoliv program Rosetta je v současnosti nabízen, nevyužívá jej podle registrů OpenDOAR a ROAR ani jedna instalace.

Na software Digitool je postaven Digitální univerzitní repozitář Univerzity Karlovy³⁸. V tomto repozitáři nalezneme digitalizované historické fondy, metodiky, institucionální časopisy UK publikované v režimu Open Access, kvalifikační práce publikované před rokem 2006 a další sbírky. Nutno zmínit, že Univerzita Karlova provozuje ještě jeden institucionální repozitář a to Digitální repozitář Univerzity Karlovy³⁹, který je určen výhradně pro kvalifikační práce a který je postaven na již dříve zmíněném systému DSpace.

2.4.5 EPrints

Dalším zástupcem softwaru s otevřeným zdrojovým kódem je EPrints. Tato aplikace je vyvíjena na Southamptonské univerzitě, kde vznikl také registr ROAR, který je na tomto software postaven. Vývoj systému EPrints započal už v roce 2000, tedy v době začínajících debat o repozitářích jejich ko-operaci. Zatím poslední verze systému je verze 3.3 vydaná v roce 2011. (University of Southampton, 2011)

Software EPrints je distribuován pod licencí GNU General Public License⁴⁰ a je napsán v programovacím jazyce Perl. Systém samotný lze nainstalovat na server s operačním systémem unixového typu (např. Linux). Vývoj softwaru je řízen již zmíněnou Southamptonskou univerzitou, nicméně zdrojové soubory jsou vystaveny na platformě GitHub, kde také probíhá komunitní vývoj. Systém je multilingvální, nicméně kvůli malému využití v České republice není přeložen do češtiny. O překlad do cílového jazyka se většinou stará cílová instituce, která pak jazykovou verzi poskytne ostatním ke stažení.

Systém EPrints nabízí následující funkce (University of Southampton, 2007):

- Správu kolekcí

³⁷ Ex Libris Ltd.: <http://www.exlibrisgroup.com/>

³⁸ Odkaz do Digitálního univerzitního repozitáře UK: <http://digitool.is.cuni.cz>

³⁹ Odkaz do Digitálního repozitáře UK: <https://dspace.cuni.cz/>

⁴⁰ GNU GENERAL PUBLIC LICENSE: <https://www.gnu.org/licenses/gpl-3.0.en.html>

- Vyhledávání
- Export záznamů v různých formátech (Dublin Core, BibTex, ...)
- Nahrávání různých formátů dokumentů
- Jednoduché rozhraní pro nahrávání dokumentů do repozitáře s napovídáním v polích
- Import záznamů z externích webových služeb (PubMed, CrossRef)

Jediným repozitářem, který podle ROAR v Česku využívá software EPrints je Knihovna Akademie Věd České republiky⁴¹. V tomto repozitáři nalezneme některé časopisy publikovány touto institucí. Nicméně jedná se spíše o archiv čísel časopisů, protože veřejné stránky těchto časopisů jsou na jiných webech.

2.4.6 Informační systém Masarykovy univerzity

Informační systém MU je provozován a vyvíjen na Fakultě informatiky Masarykovy univerzity od roku 1999. Mimo funkcionality repozitáře podporuje také studijní administrativu a e-learning. Jedná se tedy o komplexní řešení pro vzdělávací instituci. Systém již získal řadu ocenění a je nasazen i na jiných institucích kromě domovské Masarykovy univerzity. (Fakulta informatiky MU, 2015)

Informační systém je nabízen formou SaaS (Software as a service). Tedy lze si jej pronajmout na rok či více s tím, že poskytovatel se stará o technický provoz systému, databázi, ladí systém, či školí uživatele.

Informace o repozitářovém modulu Informačního systému MU nejsou k dispozici nicméně několik funkcí je patrných při procházení systému:

- Vyhledávání
- Třídění do kolekcí
- Tagování prací na základě klíčových slov a vyhledání podobných prací
- Archivaci vložených souborů (každý soubor má hash kód)
- Metadatový popis Dublin Core (ve zdrojovém kódu záznamu jsou metadata Dublin Core)
- Nahrávání různých formátů dokumentů

⁴¹ Odkaz do repozitáře: <http://dlib.lib.cas.cz>

3 Vybrané standardy a certifikace pro institucionální repozitáře

Obecně se dá říci, že aplikace auditu na repozitář má dvě roviny důsledků. Interní, kdy nemá pouze kontrolní funkci, ale také pomáhá nastavit mechanismy a procesy v repozitáři a jeho řízení tak, aby fungoval efektivněji a spolehlivěji. Dále také externí, kdy deklaruje cílové skupině repozitáře, že daný repozitář je spolehlivý. Tedy, že splňuje jasně definovaná kritéria odpovídající současným požadavkům.

V této kapitole bude shrnut historický vývoj standardů, který začíná už v roce 2002. Dále bude blíže vysvětlen termín důvěryhodný repozitář. Nicméně zejména budou popsány standardy a certifikace, které lze dnes na důvěryhodné institucionální repozitáře aplikovat. Mezi certifikacemi budou zmíněny také dvě tzv. pečeti či známky (seal) - the Nestor Seal for Trustworthy Digital Archives a Data Seal of Approval.

3.1 Historický vývoj standardů a certifikací

Jak už bylo řečeno v úvodu práce prvním krokem ke sjednocování repozitářů a jejich fungování bylo vydání dokumentu Reference Model for an Open Archival Information System (OAIS) v roce 2002. Model OAIS sjednocuje terminologii ohledně repozitářů, protože do té doby neexistovala standardizace ohledně implementace repozitáře. OAIS stanovuje obecný model repozitáře a pojmenovává jeho části.

Referenční model OAIS využívá termín digitální archiv pro pojem, který je v současné odborné literatuře chápán jako repozitář či digitální repozitář. (The Center for Research Libraries a Online Computer Library Center, 2007; Consultative Committee for Space Data Systems, 2011) viz například TRAC (2007) Zároveň v těchto dokumentech zabývajících se standardizací a certifikací repozitářů jsou pod pojmy repozitář a digitální repozitář chápány jakékoli digitální repozitáře a archivy, které nabízejí a mají odpovědnost za dlouhodobé uchovávání svého obsahu.

Ve stejném roce jako model OAIS byl pracovní skupinou organizací RLG⁴² a OCLC⁴³ vydán dokument Trusted Digital Repositories: Attributes and Responsibilities⁴⁴, který specifikuje vlastnosti a požadavky na repozitář, který by se mohl nazývat důvěryhodným, spolehlivým a

⁴² Research Libraries Group – V roce 2006 se tato organizace začlenila do OCLC viz <http://www.oclc.org/research/partnership/history.html>

⁴³ Mezinárodní nezisková organizace sdružující klíčové členy knihovnické komunity: <http://www.oclc.org>

⁴⁴ Attributes of Trusted Digital Repositories: <http://www.oclc.org/research/activities/trustedrep.html>

udržitelným, a který by byl zároveň schopný uchovávat materiály ze sbírek velkých i malých kulturních a vědeckých institucí. Jde tedy o sadu doporučení.

O rok později v roce 2003 byla organizacemi RLG a NARA⁴⁵ vytvořena pracovní skupina za účelem vytvoření metodiky pro certifikaci digitálních repozitářů. Tato metodika byla dokončena v roce 2007 a nazývá se Trustworthy Repositories Audit & Certification: Criteria and Checklist neboli TRAC. (The Center for Research Libraries a Online Computer Library Center, 2007)

Rok před dokončením nástroje TRAC, v roce 2006, byl v Německu skupinou Nestor⁴⁶ vytvořen nástroj Catalogue of Criteria for Trusted Digital Repositories⁴⁷. Nestor je pracovní skupina složená z odborníků napříč oblastmi knihovnictví, muzejnictví, ale i počítačové vědy. Nástroj má podobnou strukturu jako TRAC nicméně není tak podrobný, což lze vidět i na počtu otázek v seznamu požadavků.

V roce 2007 vzniká nástroj DRAMBORA⁴⁸, který je vytvořen na základě mnoha proběhlých auditů repozitářů. Tento nástroj není samotným auditačním dokumentem či metodikou, ale slouží k určení a řízení rizik repozitáře, která chápe jako hlavní problém v oblasti důvěryhodných repozitářů. Nástroj je k dispozici jak v offline verzi, tak i online.

Dva roky po nástroji DRAMBORA vzniká další podpůrný nástroj - PLATTER. Ten byl vytvořen v rámci Evropského projektu FP6 organizací DPE⁴⁹. Ani PLATTER není nástrojem pro audit jako takový, ale slouží k přípravě na audit. Dokument je určen pro všechny zainteresované osoby při vývoji a provozu repozitáře, zejména manažery a IT pracovníky. V roce 2010 se sešli představitelé organizací CCSDS⁵⁰, DSA⁵¹ a pracovní skupiny DIN⁵² pro certifikaci důvěryhodných repozitářů, aby podepsali tzv. Memorandum of Understanding.

⁴⁵ National Archives and Records Administration: <https://www.archives.gov/>

⁴⁶ Network of expertise in long-term storage of digital resources in Germany: http://www.langzeitarchivierung.de/Subsites/nestor/EN/Home/home_node.html

⁴⁷ Overview of current standards in digital preservation http://www.langzeitarchivierung.de/Subsites/nestor/EN/Standardisierung/standardisierung_node.html

⁴⁸ Digital Repository Audit Method Based on Risk Assessment - <http://www.repositoryaudit.eu/>

⁴⁹ DigitalPreservationEurope; V dnešní době už tato organizace nefunguje. Více informací na: <http://www.digitalpreservation.gov/series/edge/dpe.html>; Tato organizace se zasadila i o propagaci problematiky dlouhodobého uchovávání sérií animovaných klipů na platformě YouTube: <https://www.youtube.com/user/wepreserve>

⁵⁰ Consultative Committee for Space Data Systems: <https://public.ccsds.org/default.aspx>

⁵¹ Data Seal of Approval: <https://www.datasealofapproval.org/en/>

⁵² Deutsches Institut für Normung (German Institute for Standardization): <http://www.din.de/en>

Tento dokument podpořený Evropskou komisí představuje závazek všech tří stran, na společném postupu při stanovování obecného rámce pro certifikaci digitálních repozitářů. Dohodli se na třech úrovních certifikace: (CCSDS et al., 2010)

- Základní certifikace – repozitář, který získá Data Seal of Approval
- Rozšířená certifikace – repozitář, který navíc k DSA provede strukturovaný a veřejně dostupný vlastní audit zhodnocený externě a založený na ISO 16363⁵³ nebo DIN 31644⁵⁴
- Formální certifikace – repozitář, který navíc k DSA získá také plný externí audit podle ISO 16363 nebo DIN 31644

První verze postupu pro získání známky Data Seal of Approval byla vydána v roce 2009. Jako druhý dokument v pořadí v oblasti certifikace repozitářů byl vydán Audit and Certification of Trustworthy Digital Repositories, který byl vytvořen v roce 2011 mezinárodní organizací CCSDS. Tento dokument byl následně zpracován jako norma pro auditování důvěryhodných repozitářů ISO 16363:2012 (ISO 16363:2012, 2012) a lze jej využívat k profesionálním i vlastním auditům. V roce 2012 byl pak vydán poslední ze tří nástrojů pro certifikaci a to známka Nestor Seal for Trustworthy Digital Archives.

3.2 Důvěryhodný repozitář

V roce 2003 byl publikován standard ISO 14721⁵⁵ popisující prostředí, funkční komponenty a informační objekty systému pro dlouhodobé uchovávání a vycházející z modelu OAIS. Tento dokument se zasloužil o sjednocení terminologie a tedy vytvoření společné půdy pro debatu o repozitářích. Nicméně mnohé instituce a komunity začaly model využívat poněkud jinak, než bylo zamýšleno.

Aby podtrhly svou spolehlivost a důvěryhodnost, začaly instituce prohlašovat, že jsou "OAIS-compliant" (volný překlad: odpovídající OAIS). Nicméně vzhledem k tomu, že OAIS je pouze model či terminologie pro popis repozitáře a nedefinuje konkrétní implementaci repozitáře, toto prohlášení nezaručovalo důvěryhodnost a spolehlivost repozitáře. (TRAC 2007)

⁵³ ISO 16363:2012 Space data and information transfer systems -- Audit and certification of trustworthy digital repositories: <https://www.iso.org/standard/56510.html>

⁵⁴ DIN 31644: Information and documentation - Criteria for trustworthy digital archives: <http://www.din.de/en/getting-involved/standards-committees/nid/standards/wdc-beuth:din21:147058907>

⁵⁵ ISO 14721:2012 Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model: <https://www.iso.org/standard/57284.html>

Rosenthal et al. (2005) specifikuje potenciální hrozby a rizika, se kterými musí důvěryhodný digitální repozitář počítat. Patří mezi ně: selhání nosiče dat, chyba hardwaru, chyba softwaru, chyba při přenosu dat, chyba síťových služeb, zastarávání nosičů dat a hardwaru, zastarávání softwaru, chyby obsluhy, přírodní katastrofy, útoky zvenčí, útoky zevnitř repozitáře, ekonomické potíže, chyby v samotné organizaci. V auditu TRAC (2007) je dále zmíněno, že tyto hrozby a rizika půjde překonat jen při nasazení neustálého monitorování, plánování, údržby, dále také pomocí vědomých zásahů a strategie.

Důvěryhodný repozitář by tedy měl být schopen zaručit a dokázat, že tyto hrozby jsou ošetřeny a že s nimi je při návrhu repozitáře počítáno. Nicméně se to netýká pouze vhodného návrhu systému nebo výběru vhodného existujícího řešení, ale musí být na místě také jasné procesy a politiky, které ochrání data v repozitáři i pokud tyto hrozby nastanou.

3.3 Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)

Tento dokument není prvním oficiálním příspěvkem k tématu stanovení a ověření důvěryhodného repozitáře, nicméně právě na tomto dokumentu staví většina následujících auditů. Ačkoliv vznikl od roku 2003 až do roku 2007, a bylo na něj jistě vynaloženo nemálo času a prostředků, je volně dostupný na internetu⁵⁶. Audit je určen jak pro vlastní vyhodnocení repozitářů institucemi, tak i pro externí certifikační entity.

Dokument stanovuje základní definici důvěryhodného digitálního repozitáře a shrnuje komponenty, které musí být prověřeny a vyhodnoceny. Dále popisuje jednotlivé body, které musí důvěryhodný repozitář splnit. K těmto bodům je vždy uveden název, popis a způsob, jak lze splnění daného bodu doložit. Na konci dokumentu je přiložen formulář pro auditora s připravenými poli podle jednotlivých bodů auditu.

Audit je členěn do tří hlavních oblastí pokrývajících organizační infrastrukturu, správu uložených digitálních objektů a použitou technologii, technickou infrastrukturu a bezpečnost. Níže jsou vypsány jednotlivé skupiny zájmu tohoto auditu tak, jak jsou uvedeny v dokumentu. (The Center for Research Libraries a Online Computer Library Center, 2007)

A) Organizační infrastruktura (24 vyhodnocovaných bodů)

- A1) Vedení repozitáře a jeho životaschopnost
- A2) Organizační struktura a personál

⁵⁶ Trustworthy Repositories Audit & Certification: Criteria and Checklist:
<http://catalog.crl.edu/record=b2212602~S1>

- A3) Odpovědnost za procedury a politika
- A4) Finanční udržitelnost
- A5) Kontrakty, licence a závazky
- B) Správa digitálního objektu (44 vyhodnocovaných bodů)
 - B1) Přijetí: Akvizice obsahu
 - B2) Přijetí: Vytvoření Archival Information Package (AIP)
 - B3) Plán uchovávání
 - B4) Ukládání a údržba AIP (aplikace plánu uchovávání)
 - B5) Správa informací (metadata, data vznikající během uchovávání)
 - B6) Správa přístupu k AIP (tvorba DIP, nebo něco jiného)
- C) Technologie, technická infrastruktura a bezpečnost (16 vyhodnocovaných bodů)
 - C1) Infrastruktura systému (OS, SW, HW; řízení změn)
 - C2) Používání vhodných technologií (HW, SW; sledování požadavků a prostředí)
 - C3) Bezpečnost (fyzické zajištění HW, software zajištění, ale i personální)

3.4 Catalogue of Criteria for Trusted Digital Repositories

Tento dokument je chápán jako první větší oficiální dokument reagující na v té době nejasnou definici důvěryhodného repozitáře. První verze byla publikována v roce 2006⁵⁷ pracovní skupinou NestorWorking Group -Trusted Repositories Certification složenou z německých odborníků pracujících v muzeích, knihovnách a archivech, ale i vydavatelstvích, vědeckých institucích a certifikačních společnostech a ve vývoji software. V roce 2009 vyšla revidovaná verze dokumentu, která je čitelnější a přehlednější, ale vesměs nepřináší velké změny. Změnil se checklist na konci dokumentu určený pro vyhodnocení repozitáře, který je nově navržen jako seznam bodů a nikoliv jako tabulka na odškrtnutí. Dále byly upraveny některé body checklistu (Working Group on Trusted Repositories Certification, 2009):

- Přidán bod 4.6: Continuation of the preservation tasks is ensured even beyond the existence of the DR [pozn. Digital Repository].
- Odebrán bod 5.3: The DR reacts to substantial changes

Ve své struktuře je tento dokument podobný nástroji TRAC. Jeho hlavní částí je totiž seznam hodnocených kritérií, jejichž splněním lze repozitář považovat za důvěryhodný. Nástroj také nabízí jednoduchý formulář se všemi kritérii a úrovněmi jejich splnění pro zhodnocení repozitáře.

Rozdíl oproti nástroji TRAC je ve struktuře popisu jednotlivých kritérií. Každé hodnocené kritérium je popsáno svým názvem, obecným popisem, dále komentářem ve formě příkladů a poznámek k aplikaci a konečně také celkem obsáhlým seznamem literatury související

⁵⁷ Původní první vydání: http://files.dnb.de/nestor/materialien/nestor_mat_08-eng.pdf

s kritériem. Tento dokument je tedy také velmi nabitým informačním zdrojem s referencemi k tématu v německém i anglickém jazyce. Možná právě proto je tento dokument nazván katalog kritérií a nikoliv checklist.

Samotný seznam kritérií je členěn do velmi podobných kategorií jako v případě TRAC a to: Organizační rámec (21 kritérií), Řízení (digitálního) objektu (29 kritérií), Infrastruktura a bezpečnost (4 kritéria). Každé kritérium má pak pět různých stavů naplnění: základní koncept, naplánováno, implementováno, vyhodnoceno a publikována dokumentace.

3.5 Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)

Nástroj DRAMBORA vznikl v roce 2007 jako výsledek spolupráce organizací DCC⁵⁸ a DPE. Organizace DCC prováděla v roce 2006 a 2007 profesionální audity repozitářů a nástroj DRAMBORA je logickým vyústěním těchto aktivit. (The Digital Curation Center a DigitalPreservationEurope, 2015)

DRAMBORA chápe uchovávání digitálních objektů jako aktivitu řízení rizik. Nástroj je určen pro vlastní vyhodnocení personálem repozitáře nebo jeho manažery za účelem analýzy rizik pro repozitář a úpravy aktivních plánů. Díky tomu, že je nástroj šířen pod licencí Creative Commons, je jeho stažení zdarma. Nicméně je nutné vyplnit jednoduchou registraci sloužící pro evidenční potřeby tvůrců auditu.

Nástroj je možné získat jak v offline verzi, tak i online. Součástí offline verze je Microsoft Word dokument obsahující formulář pro upřesnění a specifikaci mandátu repozitáře, jeho cílů a na základě toho určení, popsání a klasifikaci rizik. Zde se zapisují výsledky auditu a zároveň tento dokument slouží jako reference pro všechny aktivní dokumenty, plány, rizika a osoby s těmito spojenými. Tento dokument je tedy výstupem samotného auditu. Dokument je přiložen také ve verzi pro program Microsoft Excel, která obsahuje makra a vzorce pro jednodušší práci s formulářem. Stěžejní součástí nástroje je obsáhlý PDF dokument popisující průběh auditu a upřeshňující a vysvětlující jednotlivé sekce auditního formuláře zmíněného výše. Online verze auditu je přístupná pouze po zaregistrování repozitáře, který projde auditem. (The Digital Curation Center a DigitalPreservationEurope, 2007)

Audit se skládá ze šesti úrovní a zaměřuje se na následující části.

1. úroveň: Identifikace organizačního kontextu
2. úroveň: Dokumentace existujících politik a regulací

⁵⁸ Digital Curation Centre: <http://www.dcc.ac.uk/>

- 3. úroveň: Identifikace činností, aktiv a jejich zodpovědných osob
- 4. úroveň: Identifikace rizik
- 5. úroveň: Ohodnocení rizik
- 6. úroveň: Řízení rizik

Každá úroveň je formalizovaně popsána podčástmi. Tyto podčásti vysvětlují, co je cílem každé úrovně, popisují diagram úrovně, co se očekává jako výsledek aj. Vyčerpávající výpis následuje.

- 1. Cíl úrovně
- 2. Úkoly spojené s touto úrovní
- 3. Očekávané výsledky úrovně
- 4. Kontext dané úrovně v rámci auditu
- 5. Jaké zdroje jsou potřeba k dokončení úrovně
- 6. Diagram znázorňující danou úroveň
- 7. Instrukce pro dokončení úrovně
- 8. Co dělat, pokud nejsou k dispozici potřebné informace
- 9. Co vyplnili jiné repozitáře
- 10. Diskuse k úrovni
- 11. Seznam nejdůležitějších úkolů v této úrovni (checklist)

Kvůli své podrobnosti je absolvování auditu DRAMBORA vhodným krokem, pokud se management repozitáře rozhodne jakýkoli oficiální audit v budoucnu podstoupit. Navíc v praxi by takový repozitář musel informace získané díky nástroji DRAMBORA shánět pro přípravu k profesionálnímu auditu i tak.

3.6 Planning Tool for Trusted Electronic Repositories (PLATTER)

Původní anglická verze PLATTER vznikla v rámci projektu organizace DPE, který byl financován v rámci šestého rámcového programu FP6. O překlad do českého jazyka se postarali Jan Hutař, Ladislav Cubr, Marek Melichar a Andrea Fojtů. Přeložený dokument vydala v roce 2009 Národní knihovna České republiky. (Rosenthal et al., 2009)

Tento dokument není auditem, ani standardem. Je to souhrnný dokument určený pro seznámení se s problematikou dlouhodobé ochrany digitálních dat psaný tak, aby mu porozuměl i laik, který se na tuto oblast nespecializuje. Cílem dokumentu je „*být prvním dokumentem, který vezmeme do ruky před tím, než začneme budovat repozitář nebo než se začneme snažit o jeho důvěryhodnost.*“ (tamtéž) Cílovou skupinou jsou mimo jiné manažeři a

vedoucí pracovníci či zaměstnanci IT oddělení. V souboru nástrojů pro zajištění důvěryhodného repozitáře tedy PLATTER zabírá první místo a měl by být konzultován ještě před tím, než se rozhodneme pro jakýkoli audit, ideálně ještě před vznikem samotného repozitáře.

První částí nástroje PLATTER je klasifikace repozitáře. Pomocí odpovídání na jednoduché otázky je uživatel schopen definovat svůj repozitář podle čtyř hlavních tříd: účel a funkce, velikost, provoz a technické řešení a možnosti implementace. Tato definice zároveň umožňuje porovnávat mezi sebou nastavení podobných repozitářů.

Druhou část PLATTERu představuje návrh plánovacího cyklu strategických cílů a poté sada předpřipravených strategických cílů. Nejdříve je popsán formalizovaný soubor kroků pro definici a specifikaci cílů organizace, jejich implementaci a kritéria jejich vyhodnocení. Těmito kroky jsou: strategické plánování, definice cílů, plánování činností, realizace, hodnocení a reformulace. V kroku strategického plánování je určen mandát repozitáře, zainteresované subjekty (stakeholders) a výkon repozitáře. V dalším kroku jsou stanoveny definice cílů podle principu SMART - konkrétní (specific), měřitelné (measurable), splnitelné (achievable), realistické (realistic) a časově vymezené (time managed). V kroku plánování činností dochází k nastavení postupů pro naplnění všech stanovených cílů nebo alespoň těch nejdůležitějších. Následuje cyklický sled kroků realizace, hodnocení a reformulace, které probíhají periodicky od spuštění repozitáře až do konce jeho provozu. Zde se na základě aktuálních dat z provozu a hodnocení splnění metrik hodnocení upravují plány činností.

Největší výhodou PLATTERu je však předpřipravený soubor plánů pro strategické cíle. Tedy vlastně již předpřipravené první tři kroky plánovacího cyklu. Těchto plánů je celkem devět a vycházejí z Deseti základních principů důvěryhodného repozitáře⁵⁹. Každý z plánů obsahuje vyjmenované důležité cíle. Každý z těchto cílů má pak uvedeny příklady implementace a komentář popisující kontext cíle. Níže je seznam plánů.

1. Finanční plán
2. Plán řízení lidských zdrojů
3. Datový plán

⁵⁹ THE DIGITAL CURATION CENTER, UK, DIGITALPRESERVATIONEUROPE, NESTOR (GER) a THE CENTER FOR RESEARCH LIBRARIES, 2007. Ten Principles. *Archiving & Preservation* [online] [vid. 2017-05-13]. Dostupné z: <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re>

4. Akviziční plán
5. Plán zpřístupňování
6. Plán ochrany
7. Technický plán
8. Plán zajištění kontinuity
9. Krizový plán

Nástroj PLATTER není samostatným nástrojem, který by sám sloužil ke stanovení důvěryhodnosti repozitáře. Avšak tvůrci PLATTERu tvrdí, že: „*Repozitář, který přistoupí na filozofii PLATTERu a bude se věnovat všem bodům obsaženým v rozboru plánů strategických cílů (...), si může být jistý, že takto zahrnul do svého plánování všechny hlavní body z TRAC i Nestor,*“ Zároveň uvádějí, že je PLATTER komplementárním nástrojem pro analýzu rizik DRAMBORA.

3.7 Data Seal of Approval

Certifikace DSA spadá do kategorie tzv. vlastních certifikací, kdy instituce spravující repozitář sama provede audit podle předem daných kritérií a požadavků. Dle Evropského rámce pro audit a certifikaci digitálních repozitářů⁶⁰ je to nejnižší úroveň certifikace, tzv. základní certifikace. Výsledek auditu spolu s příslušnými dokumenty pak instituce zašle k posouzení organizaci spravující certifikaci. Poté je případně vydána repozitáři certifikace Data Seal of Approval spolu s možností zveřejnit toto ocenění na stránkách repozitáře umístěním loga. Samotný vlastní audit a posudek repozitáře jsou po udělení certifikace veřejně dostupné⁶¹. Certifikace se uděluje na tři roky a je nutno ji tedy obnovovat. (Data Archiving and Networked Services, 2017)

Časové omezení platnosti má pozitivní vliv jak na instituci, která musí periodicky přemýšlet nad nastavením svého repozitáře a politik, tak i na samotnou certifikaci, která může reagovat na změny a přizpůsobit certifikaci podle aktuálních potřeb komunity pro dlouhodobou archivaci. DSA je totiž dnes nejjednodušší možností certifikace repozitáře a jako taková

⁶⁰ European Framework for Audit and Certification of Digital Repositories:
<http://www.trusteddigitalrepository.eu/Memorandum%20of%20Understanding.html>

⁶¹ List of Repositories that have acquired the Data Seal of Approval:
<https://assessment.datasealofapproval.org/seals/>

začíná být vyžadována mnoha grantovými organizacemi či mezinárodními konsorcií, například Mezinárodním konsorciem sociálněvědních datových archivů CESSDA⁶².

Aktuální metodika pro hodnocení repozitáře je pro roky 2017-2019⁶³ a popis metodiky bude tedy vycházet z ní. Jazykem pro vyhodnocení certifikace je angličtina. Pokud instituce nemá k dispozici anglické verze svých interních dokumentů, je nutné přiložit shrnutí těchto dokumentů v angličtině. Celá metodika se skládá z následujících 16 oblastí požadavků:

1. Kontext (Context)
2. Mise/Mandát (Mission/Scope)
3. Licence (Licenses)
4. Udržení přístupu (Continuity of access)
5. Důvěrnost / etika (Confidentiality / Ethics)
6. Organizační infrastruktura (Organizational infrastructure)
7. Expertní dohled (Expert guidance)
8. Integrita dat a autenticita (Data integrity and authenticity)
9. Posuzování dat (Appraisal)
10. Zdokumentované procedury pro ukládání (Documented storage procedures)
11. Plán dlouhodobého uchovávání (Preservation plan)
12. Kvalita dat (Data quality)
13. Procesy (Workflows)
14. Vyhledávání a identifikace dat (Data discovery and identification)
15. Užití dat (Data reuse)
16. Technická infrastruktura (Technical infrastructure)
17. Bezpečnost (Security)

Každá oblast auditu vyžaduje vyplnění dvou polí. Prvním polem je tzv. úroveň splnění (Compliance level), kde je k dispozici 5 úrovní:

- 0 – netýká se našeho repozitáře
 - Může se stát, že zmíněná oblast pro daný repozitář neplatí. Poté musí repozitář vysvětlit a doložit, proč je daná oblast nerelevantní.
- 1 – repozitář se tímto ještě nezabýval

⁶² CESSDA prepared for the new requirements for trustworthy data repositories:
<https://www.cessda.eu/eng/Consortium/Communication/News/CESSDA/CESSDA-prepared-for-the-new-requirements-for-trustworthy-data-repositories>

⁶³ The Guideline Assessment Documentation: <https://assessment.datasealofapproval.org/documentation/>

- 2 – repozitář má teoretický návrh, jak oblast řešit
- 3 – repozitář tuto oblast momentálně implementuje
- 4 – tato oblast byla plně zapracována do repozitáře

Úrovně 0-2 jsou spíše informativní pro zaměstnance repozitáře, protože repozitář může být certifikován, až pokud splňuje u všech oblastí alespoň úroveň 3. Samozřejmě ideální stav je mít všechny oblasti na úrovni 4.

Druhým povinným polem u každého požadavku je popis, jak je daný požadavek implementován. Nicméně hlavní částí tohoto pole by měly být odkazy na příslušné politiky a směrnice dokládající implementaci.

3.8 Nestor Seal for Trustworthy Digital Archives

Tato certifikace spadá do druhé kategorie Evropského rámce pro audit a certifikaci digitálních repozitářů, tzv. Rozšířená certifikace. Metodika certifikace je postavena na německém standardu DIN 31644.

Postup pro získání této certifikace je podobný jako v případě dříve zmíněné certifikace DSA. Repozitář ustanoví kontaktní osoby a zažádá o započetí řízení o certifikaci, kdy je mu přidělen hodnotitel z certifikační autority. Poté repozitář provede vlastní audit dle poskytnutých dokumentů a tento audit následně odešle hodnotiteli. Ten vyhodnotí zaslaný audit, přepoše jej druhému hodnotiteli a tyto posudky jsou pak vráceny repozitáři a certifikační autoritě. Repozitář se pak může proti rozhodnutí hodnotitelů odvolat. Následně je v případně kladného hodnocení udělena certifikace, která však nabude platnosti až po zveřejnění samotného auditu, reportu recenzí hodnotitelů a dalších souvisejících dokumentů viditelně na webu repozitáře spolu s logem udělení certifikace. Udělená certifikace je formálně platná po neomezenou dobu. Nicméně tím, že ve zmíněném logu musí být uveden rok udělení certifikace, certifikační autorita předpokládá, že repozitáře budou chtít udržovat svou certifikaci aktuální a budou ji pravidelně obnovovat. (Nestor Certification Working Group, 2013)

V Tabulka 1 je vypsáno všech 34 kritérií certifikace Nestor. Každé kritérium musí v auditu obsahovat jednak formální bodové ohodnocení jeho naplnění a jednak popis, jak je toto kritérium naplněno. Tento popis by měl obsahovat také odkazy na příslušné směrnice či metodiky repozitáře, které dané kritérium implementují. Kritéria C1-C12 musí mít bodové ohodnocení 10 a musí být tedy plně implementována. Průměr ze součtu kritérií C13-C34 musí být větší nebo roven 7, zároveň každé kritérium z tohoto rozsahu musí mít hodnocení alespoň 6. Některá z těchto kritérií tedy mohou být ve fázi detailně naplánována.

Jednotlivá kritéria lze obodovat známkami 0, 3, 6 a 10 podle následujícího popisu:

- 0 – ještě nebylo řešeno
- 3 – je naplánováno
- 6 – je detailně naplánováno
- 10 – je implementováno do provozu

Příručka pro vlastní archivaci u každého kritéria specifikuje minimální ohodnocení, vysvětlení kritéria a jeho důležitosti, otázky, které by měly být zodpovězeny v popisu a příklady dokumentů, které mohou kritérium implementovat

Kód	Název kritéria	Min. body
C1	Výběr informačních objektů a jejich reprezentace (mise repozitáře)	10
C2	Odpovědnost za dlouhodobé uchovávání obsahu	10
C3	Cílové skupiny uživatelů	10
C4	Přístup k repozitáři	10
C5	Dlouhodobá čitelnost uchovaných dat	10
C6	Smluvní podklady	10
C7	Soulad s fungujícím právním prostředím	10
C8	Financování	10
C9	Personál	10
C10	Organizace a procesy	10
C11	Opatření pro dlouhodobé uchovávání	10
C12	Krizové řízení a předání dat	10
C13	Klíčové vlastnosti uchovávaných dokumentů	6
C14	Integrita: Rozhraní příjmu obsahu	6
C15	Integrita: Funkce k uchovávání obsahu	6
C16	Integrita: Uživatelské rozhraní	6
C17	Autenticita: Příjem obsahu	6
C18	Autenticita: Opatření k uchovávání obsahu	6
C19	Autenticita: Zpřístupnění obsahu	6
C20	Technická kontrola nad obsahem	6
C21	Specifikace nahrávaného obsahu	6
C22	Transformace nahrávaného obsahu na archivační objekt	6
C23	Specifikace archivačního objektu	6
C24	Interpretabilita archivačního objektu	6
C25	Transformace archivačního objektu na zpřístupňovaný objekt	6
C26	Specifikace zpřístupňovaného objektu	6
C27	Identifikace	6
C28	Popisná metadata	6
C29	Strukturální metadata	6
C30	Technická metadata	6
C31	Zaznamenávání opatření k dlouhodobému uchovávání	6

C32	Administrativní metadata	6
C33	IT infrastruktura	6
C34	Bezpečnost	6

Tabulka 1: Seznam kritérií certifikace Nestor Seal for Trustworthy Digital Archives a jejich minimální bodové ohodnocení

Certifikace Nestor je skutečně rozsáhlejší možností certifikace a nutně ji tedy využije méně institucí. V době psaní této práce splnily tuto certifikaci dle seznamu na stránkách certifikační autority⁶⁴ dvě instituce. Je to sice malé číslo pro udání obecného trendu, ale vzhledem k tomu, že obě instituce jsou národními repozitáři, autor práce odhaduje, že po této certifikaci sáhnou spíše velké národní či mezinárodní repozitáře.

3.9 Audit and Certification of Trustworthy Digital Repositories

Manuál popisující doporučený postup pro provádění auditů repozitářů. Tento dokument vychází a staví na nástroji TRAC. Jako základ pro terminologii v dokumentu je využita publikace Reference Model for an Open Archival Information System. Tento manuál publikovala v roce 2011 mezinárodní organizace Consultative Committee for Space Data Systems, která stojí i za zmíněným modelem OAIS. Je také nutné zmínit, že tento manuál je v podstatě rukopisem pro standard ISO 16363:2012. V Německu mají obdobu standardu ISO 16363 s názvem DIN 31644. Nicméně tato práce se zabývá pouze zdroji v anglickém či českém jazyce.

Provedení externího auditu repozitáře podle tohoto standardu je dle Evropského rámce pro audit a certifikaci digitálních repozitářů tzv. Formální certifikace. Tedy třetí a nejvyšší stupeň certifikace.

Dokument přesně specifikuje body, které musí repozitář splňovat, aby byl podle tohoto auditu certifikován jako důvěryhodný. Nicméně oproti TRAC, ze kterého vychází, nabízí u každého bodu navíc také diskusi. Tedy popis kontextu daného bodu a proč je pro audit důležitý. Tento audit je tedy více informativní než TRAC. Jednotlivé body mají strukturu: název, krátký popis, způsoby, jak splnit daný bod, a diskusi k danému bodu. Dokument neobsahuje přiložen žádný vyhodnocovací formulář tak jako TRAC. Dokument také nestanovuje, jak přesně daných bodů docílit a je na auditorovi, zda uzná řešení daného repozitáře za odpovídající.

Níže jsou vypsány jednotlivé oblasti zájmu auditu. Jak je vidět, struktura oblastí auditu se až na výjimky shoduje se strukturou TRAC.

⁶⁴ NESTOR SEAL FOR TRUSTWORTHY DIGITAL ARCHIVES - EVALUATED ARCHIVES:
<http://www.dnb.de/Subsites/nestor/EN/Siegel/siegel.html>

A) Organizační infrastruktura (21 vyhodnocovaných bodů)

- A1) Vedení repozitáře a jeho životaschopnost
- A2) Organizační struktura a personál
- A3) Odpovědnost za procedury a rámec pro politiku uchovávání obsahu
- A4) Finanční udržitelnost
- A5) Kontrakty, licence a závazky

B) Správa digitálního objektu (45 vyhodnocovaných bodů)

- B1) Přijetí: Akvizice obsahu
- B2) Přijetí: Vytvoření Archival Information Package (AIP)
- B3) Plán uchovávání
- B4) Uchovávání AIP
- B5) Správa informací (metadata, data vznikající během uchovávání)
- B6) Správa přístupu (tvorba DIP, nebo něco jiného)

C) Infrastruktura a řízení bezpečnostních rizik (19 vyhodnocovaných bodů)

- C1) Řízení rizik technické infrastruktury
- C2) Řízení bezpečnostních rizik

4 Politiky institucionálních repozitářů v ČR

Jak bylo vysvětleno v úvodu této práce, repozitář není jen hardware a software pro ukládání a zpřístupňování. Pokud má být repozitář důvěryhodný, je nutné, aby měl mandát od univerzitních orgánů pro své fungování. Zároveň by měl mít také jasně definováno, kdo je jeho cílová skupina a jaké služby pro tuto skupinu zajišťuje. V těchto parametrech se institucionální repozitáře liší. Některé instituce mohou provozovat dva repozitáře – jeden na kvalifikační práce, druhý na vědecký výstup zaměstnanců. Jiná instituce se může rozhodnout sloučit tyto dokumenty v jednom repozitáři a rozlišit je pouze pomocí kolekcí.

V této kapitole budou tedy shrnuty politiky fungování, které jsou nastavené u repozitářů institucí v České republice. V potaz budou opět brány pouze instituce, které jsou uvedené v některém z registrů repozitářů – OpenDOAR a ROAR. Registry OpenDOAR i ROAR poskytují prostor pro správce repozitářů, aby uvedli svou politiku a další faktické informace jako cílová skupina, podporované dokumenty atp. Nicméně kvalita těchto záznamů není příliš konzistentní. Proto pro určení politik jednotlivých repozitářů bude čerpáno přímo ze stránek repozitářů, příp. zřizujících institucí. Tedy, pokud nebudou informace o politikách, licenčních smlouvách s uživateli, či auditech dostupné na webových prezentacích, bude předpokládáno, že neexistují. V Tabulka 2 jsou shrnuty politiky, obsah a informace o aktuálních certifikacích zkoumaných repozitářů.

Repozitář (Typ)	Politiky	Typ obsahu	Licenční smlouva s uživatelem	Audit
LINDAT/CLARIN (oborový)	Politika nakládání s osobními údaji, metadatová politika, politika uchovávání obsahu	Data a nástroje související s oborem (korpusy, slovníky, parsery,...)	Ano; přes 50 licencí pro užití nahraného obsahu	Data Seal of Approval 2014 - 2017
Digitální repozitář UK (institucionální)	Čl. 12 Studijního a zkušebního řádu, čl. 7 Rigorózního řádu, Opatření rektora č. 13/2017	Kvalifikační práce, habilitační práce	§ 47b od. 3 zákona č. 111/1998 Sb., o vysokých školách	-
Digitální univerzitní repozitář UK (institucionální)	Metadatová politika; politika uchovávání obsahu	Archivní dokumenty, publikace, historické dokumenty, mapy	-	Mapová sbírka - Data Seal of Approval 2014 – 2017; repozitář jako celek žádná
Česká matematická digitální knihovna (oborový)	Politika užití obsahu; politika kvality obsahu	Periodika, sborníky a kapitoly knih	Smlouva s autorem / vydavatelem	-

Digitální knihovna VUT (institucionální)	Směrnice rektora č. 38/2017	Kvalifikační práce, sborníky, časopisy, publikační činnost, archivní dokumenty univerzity	§ 47b od. 3 zákona č. 111/1998 Sb., o vysokých školách; Creative Commons pro publikační činnost	-
Digitální knihovna ČVUT (institucionální)	Politika příjmu obsahu	Kvalifikační práce, publikační výstupy, univerzitní dokumenty	§ 47b od. 3 zákona č. 111/1998 Sb., o vysokých školách	-
Digitální knihovna FF MU (institucionální)	-	Odborná periodika, monografie a sborníky	-	-
Univerzitní repozitář MU (institucionální)	Směrnice MU č.4/2013 - Repozitář zaměstnaneckých děl	Publikační činnost zaměstnanců MU	Směrnice MU č.4/2013 - Repozitář zaměstnaneckých děl; Creative Commons	-
Digitální knihovna UPa (institucionální)	Směrnice č. 9/2012 – kvalifikační práce; Směrnice č. 2/2015 – publikační činnost	Kvalifikační práce, publikační činnost, odborná periodika, monografie, skripta	Creative Commons	-
Digitální repozitář VŠB-TUO (institucionální)	V kompetenci Ústřední knihovny (dle Studijního a zkušebního řádu pro studium v bakalářských, a magisterských studijních programech)	Kvalifikační práce, sborníky, časopisy, publikační činnost z WoS	Kvalifikační práce - Přístup externích znemožněn; publikační činnost – přejímá licenci od vydavatele	-
Repozitář publikační činnosti UTB (institucionální)	Rozhodnutí rektora č. 10/2015	Publikační činnost	-	-
Národní úložiště šedé literatury (národní)	Politika příjmu obsahu, politika metadat	Specifikováno v textu	Smlouva s uložitelem, Creative Commons nebo vlastní licence	DRAMBORA
Studijní knihovna pro výzkum vývoje NTK (institucionální)	Směrnice č. 1/2011	Šedá literatura zaměstnanců NTK a partnerů	Creative Commons	-

Repozitář Akademie Věd ČR (institucionální)	Metodika ukládání; Interní směrnice Akademické rady č. 4/2013; politika otevřeného přístupu v AV ČR	Publikační činnost, datové soubory	Rámcová licenční smlouva s ústavem; Autorský zákon č. 121/2000 Sb. (pro vlastní užití)	-
Digitální knihovna Západočeské univerzity v Plzni (institucionální)	Kvalifikační práce - směrnice rektora č. 26R/2013	Kvalifikační práce, publikační činnost	Autorský zákon č. 121/2000 Sb.; licence vydavatele textu	-

Tabulka 2: Shrnutí politik institucionálních repozitářů v České republice

* Více informací v textu popisujícím repozitář

4.1 Repozitář LINDAT/CLARIN

Provozovatelem tohoto repozitáře je Ústav formální a aplikované lingvistiky na Karlově univerzitě⁶⁵. Typově je repozitář oborový a přijímá dokumenty z různých institucí. Při vkládání nového záznamu do repozitáře musí uživatel nejprve souhlasit s licenčními podmínkami, kde mimo jiné také stvrzuje, že má právo vystavit nahrávaný dokument v repozitáři. Tímto jsou tedy ošetřeny případné spory ohledně autorství a práv k obsahu, protože toto břemeno je přeneseno na ukládajícího uživatele. Dále v procesu ukládání si uživatel vybere jednu z mnoha nabízených licencí užití pro vkládaný dokument, kdy má na výběr různé varianty Creative Commons, GNU licenci, MIT licenci, ale i specifické licence pro obor tohoto repozitáře. LINDAT/CLARIN veřejně vystavuje politiky, kterými se řídí. Jsou to politika nakládání s osobními údaji, metadatová politika a politika uchovávání obsahu. Repozitář nabízí návštěvníkům sekci FAQ, ve které blíže odpovídá na konkrétní možné dotazy uživatelů a také vysvětluje proč vlastně data do repozitáře nahrávat. (UFAL MFF UK, 2017)

4.2 Digitální repozitář Univerzity Karlovy

Tento repozitář provozuje Ústřední knihovna Univerzity Karlovy⁶⁶. Je to institucionální repozitář určený výhradně pro ukládání a zveřejňování kvalifikačních a habilitačních prací. Veřejně přístupná bez registrace jsou v repozitáři bibliografická data, plný text práce, přílohy, posudek oponenta a záznam o průběhu obhajoby. Ke zveřejnění je dle čl. 7 Opatření rektora č. 13/2017⁶⁷ přistoupeno nejpozději 21 dní po úspěšné či neúspěšné obhajobě. Ve výjimečných

⁶⁵ <http://ufal.mff.cuni.cz/>

⁶⁶ <https://knihovna.cuni.cz/>

⁶⁷ Zpřístupnění elektronické databáze závěrečných prací, 2017. In: Opatření rektora. Praha: Univerzita Karlova, 13/2017, čl. 7. Dostupné také z: <http://www.cuni.cz/UK-8092.html>

případech je možné si zažádat o odložení zveřejnění práce až po dobu tří let. Z důvodu zajištění kvalitní archivace je každý student povinen při nahrávání své práce do univerzitního informačního systému nahrávat práci ve formátu PDF/A-1a nebo PDF/A-2u. Jsou to speciální verze formátu PDF uzpůsobené k dlouhodobé archivaci.

Samotný repozitář je díky modernímu uživatelskému rozhraní aplikace DSpace velmi přehledný. Jednotlivé fakulty zde mají vlastní kolekce, které se dále dělí na habilitační práce, kvalifikační práce a kvalifikační práce do roku 2006. Na úrovni fakult lze dále vyfiltrovat jednotlivé součásti fakulty, práce podle hodnocení obhajoby nebo práce konkrétních vedoucích. Jedním z cílů repozitáře je také „*dlouhodobé uložení a ochrana podle národních či mezinárodních standardů a ,best practices*“,“ nicméně dle dostupných informací neproběhl žádný oficiální audit. (Knihovna UK v Praze, 2017)

Na svých webových stránkách vyjadřuje UK plán na další rozvoj repozitářů: „*Do budoucna plánujeme sjednotit zpřístupňování kvalifikačních prací pouze v rámci Digitálního repozitáře UK (provozovaného v systému DSpace) a vytvoření samostatného Repozitáře publikační činnosti.*“ (tamtéž)

4.3 Digitální univerzitní repozitář Univerzity Karlovy

Druhý repozitář Univerzity Karlovy⁶⁸ je provozován na systému od firmy Ex Libris – Digitool. O jeho provoz se také stará Ústřední knihovna Univerzity Karlovy jako v případě Digitálního repozitáře. V repozitáři nalezneme mnoho rozličných sbírek. Jednak se v něm nachází historické kvalifikační práce vydané před rokem 2006. Dále obsahuje historické sbírky, mapovou sbírku Přírodovědecké fakulty, která získala certifikaci Data Seal of Approval, digitalizované dokumenty vydané na UK, Open Access monografie nakladatelství Karolinum, Open Access články a disertace z oboru ekonomických věd institutu CERGE-EI, ale i certifikované metodiky UK. Jak je vidět z tohoto krátkého přehledu, záběr tohoto repozitáře je velmi široký. Repozitář nezveřejňuje politiky související s jeho provozem či ukládáním.

4.4 Česká matematická digitální knihovna

Tento oborový repozitář slouží především pro české publikace a monografie v oboru matematiky. Byl vytvořen jako výsledek projektu 1ET200190513⁶⁹ financovaného Akademií

⁶⁸ Digitální univerzitní repozitář - <http://digitool.is.cuni.cz>

⁶⁹ Project 1ET200190513: <http://project.dml.cz/>

věd České republiky. O provoz repozitáře se stará Institut matematiky AV ČR⁷⁰. Každá přijatá publikace musí být doplněna smlouvou s autorem, resp. s vydavatelem v případě celých časopisů či sborníků. Vzhledem k tomu, že repozitář se zajímá o matematickou literaturu od 19. století, tvoří zpřístupňovaný obsah PDF soubory získané buď digitalizací starých publikací, nebo u novějšího obsahu přímo digitální verze. Digitalizované dokumenty mají stanovenou úroveň kvality. Zpřístupnění obsahu uživateli se řídí podmínkami užití⁷¹, jejichž souhlas je podmínkou pro zobrazení či stažení jakéhokoli obsahu z repozitáře. Repozitář využívá starší verzi softwaru DSpace.

4.5 Digitální knihovna Vysokého učení technického

Tento institucionální repozitář provozuje Ústřední knihovna VUT⁷² a je postaven na systému DSpace. Uvnitř jsou archivovány a zpřístupňovány jak klasifikační práce, tak i publikační činnost zaměstnanců VUT. Dále zde nalezneme sborníky z konferencí pořádaných na VUT, archivní čísla časopisů publikovaných na VUT a další dokumenty univerzity jako například výroční zprávy. Zveřejňování kvalifikačních prací v repozitáři se řídí Směrnicí č. 38/2017⁷³, kde se uvádí, že: „*Veřejná VŠKP je zveřejněna v Digitální knihovně.*“ Po otevření záznamu kvalifikační práce je k dispozici bez registrace jak plný text práce, tak i posudky oponenta a vedoucího práce a výsledná známka z obhajoby.

Repozitář nemá viditelně nastaveny konkrétní politiky, ale příjem obsahu je řízen. Kvalifikační práce jsou automaticky stahovány z univerzitního informačního systému. Autoarchivace vlastních prací vědci probíhá přes rozhraní aplikace Apollo, což je modul univerzitního informačního systému. Vědec, který se rozhodne svou práci zveřejnit v repozitáři, vyplňuje přímo v univerzitním informačním systému základní metadata jako typ licence (Creative Commons), typ nahraného souboru (preprint, postprint, vydavatelská verze), souhlas spoluautorů, datum zveřejnění (např. pro potřeby embarga vydavatele) a zda se jedná o recenzovanou práci. Tato data spolu se souborem se pak dostanou k zaměstnanci repozitáře, který je ověří. (ÚSTŘEDNÍ KNIHOVNA VUT V BRNĚ, 2014) Na webové prezentaci

⁷⁰ Institute of Mathematics of the CAS: <http://www.math.cas.cz/index.php?lang=1>

⁷¹ Conditions of Use: <http://dml.cz/use>

⁷² Ústřední knihovna VUT: <https://www.vutbr.cz/uk>

⁷³ Směrnice č. 37/2017 - ÚPRAVA, ODEVZDÁVÁNÍ, ZVEŘEJŇOVÁNÍ A UCHOVÁVÁNÍ VYSOKOŠKOLSKÝCH KVALIFIKAČNÍCH PRACÍ - <https://www.vutbr.cz/uredni-deska/vnitni-predpisy-a-dokumenty/smernice-c-38-2017-uprava-odevzdavani-zverejnovani-a-uchovavani-vysokoskolskych-kvalifikacnich-praci-d143585>

repozitáře ani Ústřední knihovny nebyly nalezeny žádné konkrétní závazky či politika týkající se dlouhodobé archivace. Zároveň ani informace, zda proběhl audit repozitáře.

4.6 Digitální knihovna Českého vysokého učení technického

České vysoké učení technické má jediný institucionální repozitář a to Digitální knihovnu ČVUT. Tento repozitář je postaven na systému DSpace a spravuje jej Ústřední knihovna ČVUT⁷⁴ ve spolupráci s Výpočetním a informačním centrem ČVUT⁷⁵. V repozitáři jsou dostupné kvalifikační práce, publikační činnost ak. pracovníků, archivní čísla vědeckého časopisu a další univerzitní dokumenty. Vkládání kvalifikačních prací do repozitáře se řídí Studijním a zkušebním řádem⁷⁶, samotné uložení probíhá automaticky skrze univerzitní informační systém. Přes rozhraní repozitáře je bez jakýchkoli zábran přístupný jak plný text práce, tak i posudky oponentů. Vkládání publikační činnosti se děje také automatizovaně z aplikace V3S⁷⁷ vyvíjené na ČVUT. (Ústřední knihovna ČVUT, 2017)

Příjem obsahu probíhá automaticky, nicméně je možné výjimečně vložit také jednotlivý záznam, k čemuž musí uživatel kontaktovat správce repozitáře. (tamtéž) Kvalifikační práce v repozitáři mají vlastní licenci pro zpřístupnění v rámci zákona. Publikační činnost nemá licenci využívání buď stanovenu vůbec, nebo přejímá licenci od vydavatele. Repozitář nemá nikde stanoven závazek k dlouhodobé archivaci obsahu provedení jakéhokoli auditu.

4.7 Digitální knihovna Filozofické fakulty Masarykovy univerzity

O správu tohoto repozitáře se stará Centrum inovačních technologií FF MU⁷⁸ a Knihovnicko-informační centrum MU⁷⁹. Svým zpracováním a použitým systémem DSpace vychází z již dříve zmíněné České matematické digitální knihovny. Cílem tohoto institucionálního repozitáře je zpřístupnění odborné produkce Filozofické fakulty, pokud možno v režimu otevřeného přístupu. Proto zde nalezneme odborná periodika, monografie a sborníky prací. (Filozofická fakulta MU, 2015)

Repozitář nemá veřejně vystaveny politiky jeho fungování. Podle jednotlivých záznamů lze však usuzovat, že vlastníci autorských práv, nejčastěji jednotlivé ústavy fakulty, se mohou

⁷⁴ Ústřední knihovna ČVUT: <http://knihovna.cvut.cz/>

⁷⁵ Výpočetní a informační centrum ČVUT: <http://vic.cvut.cz/>

⁷⁶ Studijní a zkušební řád pro studenty ČVUT ze dne 8. 7. 2015: <https://www.cvut.cz/vnitri-predpisy>

⁷⁷ Výsledky vědy a výzkumu: <https://v3s.cvut.cz>

⁷⁸ Centrum inovačních technologií FF MU: <http://www.phil.muni.cz/wcit>

⁷⁹ Knihovnicko-informační centrum MU: <http://www.muni.cz/services/library/kic-mu>

rozhodnout, zda zpřístupní dokumenty v režimu otevřeného přístupu, s embargem či vůbec. U jednotlivých dokumentů není jasně stanovena licence užívání. Zároveň není jasný závazek repozitáře k dlouhodobé archivaci, zda byly provedené nějaké kroky vstříc její realizaci, ani zda byl repozitář podroben nějakému auditu.

4.8 Univerzitní repozitář Masarykovy univerzity

Univerzitní repozitář MU je příkladem vlastního řešení. Je totiž postaven jako modul informačního systému univerzity. Repozitář má jediný účel a to shromažďování a archivaci zaměstnaneckých děl MU. Systém spravuje FI MU⁸⁰. Díky těsnému navázání repozitáře na univerzitní informační systém lze jeho obsah třídit dle kritérií jako obor či organizační jednotka. Fungování repozitáře a jeho politiky zřejmě upravuje Směrnice MU č.4/2013⁸¹, nicméně ta není dostupná pro externí uživatele. Pro stanovení možností užití zveřejněných prací jsou dostupné licence Creative Commons. Repozitář uvádí na své webové prezentaci závazek k archivaci zaměstnaneckých děl, nicméně není zmíněno provedení jakéhokoliv auditu. (FI MU, 2017)

Univerzita provozuje separátní repozitář pro kvalifikační práce, který je také implementován jako modul Thesis⁸² v univerzitním informačním systému.

4.9 Digitální knihovna Univerzity Pardubice

Tento repozitář je postaven na řešení DSpace a je ve správě Univerzitní knihovny⁸³. Uvnitř repozitáře nalezneme všechny relevantní typy dokumentů – Kvalifikační práce, publikační činnost zaměstnanců univerzity, odborné časopisy, sborníky z konferencí, knihy a skripta. Způsob ukládání kvalifikačních prací do repozitáře stanovuje Směrnice č. 9/2012⁸⁴. Ve směrnici je také podrobně popsána zodpovědnost jednotlivých aktérů (uložitel, ústav, vedoucí,...) za konkrétní úkony spojené s uložením práce – tedy jednak, kdo má daný úkon provést a jednak, kdo má provedení daného úkonu zkontrolovat. **Takto podrobný popis je v prostředí českých institucionálních repozitářů čestnou výjimkou.**

⁸⁰ Fakulta informatiky Masarykovy univerzity: <http://www.fi.muni.cz/>

⁸¹ Směrnice MU č.4/2013 - Repozitář zaměstnaneckých děl: https://is.muni.cz/do/mu/Uredni_deska/Predpisy_MU/Masarykova_univerzita/Smernice_MU/

⁸² Absolventi a závěrečné práce: <https://is.muni.cz/thesis/>

⁸³ Univerzitní knihovna Univerzity Pardubice: <http://www.upce.cz/knihovna/index.html>

⁸⁴ Směrnice č. 9/2012: <http://www.upce.cz/ff/kaa/info-bp-dp/smernice.pdf>

Ukládání a zveřejňování publikační činnosti zaměstnanců UPa upravuje Směrnice č. 2/2015. Bohužel tato směrnice není veřejně dostupná. Nicméně veřejně dostupné jsou zdroje vytvořené právě Univerzitní knihovnou⁸⁵. Dále na webu knihovny v sekci určené otevřenému přístupu lze zjistit, že: (Univerzitní knihovna Univerzity Pardubice, 2017)

„Na UPa jsou autoři na základě Směrnice č. 2/2015 povinni u výsledků typu J (článek v odporném[sic!] periodiku) a D (článek ve sborníku) přiložit k záznamu v OBD plný text příspěvku v podobě tzv. postprintu (accepted version, accepted manuscript). Plný text bude následně dle pravidel vydavatele zveřejněn v Digitální knihovně UPa.“

Na webové prezentaci repozitáře ani Univerzitní knihovny nebyl nalezen závazek k dlouhodobé archivaci uložených dokumentů, ani provedení auditu.

4.10 Digitální repozitář Vysoké školy báňské - Technické univerzity Ostrava

Repozitář VŠB-TUO spravuje Ústřední knihovna univerzity⁸⁶. Systém je postaven na software DSpace ve verzi 5.5. Obsah repozitáře tvoří kvalifikační práce, články ze sborníků a časopisů univerzity a publikační činnost akademických pracovníků zaznamenaná v citační databázi Web of Science. Ukládání kvalifikačních prací je dle Studijního a zkušebního řádu pro studium v bakalářských⁸⁷ a magisterských⁸⁸ studijních programech „v kompetenci *Ústřední knihovny za podpory informačního systému*“. Kvalifikační práce uložené v repozitáři jsou online přístupné pouze studentům či zaměstnancům VŠB-TUO a nejsou veřejně online dostupné. Jediná možnost, jak se ke kvalifikačním pracím mohou dostat externí uživatelé je osobní návštěva knihovny a registrace. (Ústřední knihovna VŠB-TUO, 2017)

Pro sběr publikační činnosti není veřejně vystavena žádná metodika a ani ve směrnicih univerzity není k dohledání směrnice regulující ukládání do repozitáře. Na rozcestníku v repozitáři věnovanému publikační činnosti se píše: „*Kolekce obsahuje bibliografické záznamy publikační činnosti (článků) akademických pracovníků VŠB-TUO v časopisech (a v Lecture Notes in Computer Science) registrovaných ve Web of Science od roku 1990 po současnost.*“ (VŠB-TUO, 2016) Vzhledem k těmto skutečnostem autor práce usuzuje, že

⁸⁵ Evidence publikační činnosti dle Směrnice č. 2/2015: se zaměřením na ukládání plných textů: <http://hdl.handle.net/10195/61947>

⁸⁶ Ústřední knihovna VŠB-TUO: <http://knihovna.vsb.cz/>

⁸⁷ Studijní a zkušební řád pro studium v bakalářských studijních programech VŠB-TUO: <https://www.vsb.cz/docs/files/cs/5b6aa096-3170-4a7f-8428-bb0c18efd06d>

⁸⁸ Studijní a zkušební řád pro studium v magisterských studijních programech VŠB-TUO: <https://www.vsb.cz/docs/files/cs/8841d5fa-978f-4642-91bf-f766f82beea6>

repozitář je plněn výhradně daty z citační databáze Web of Science a samotná univerzita, ani autoři se na plnění repozitáře bezprostředně nepodílí. Na druhou stranu v repozitáři lze nalézt několik stovek plných textů prací zpřístupněných v režimu otevřeného přístupu. Tyto práce jsou tam pravděpodobně nahrány s úsilím knihovníků spravujících repozitář. Licence k užívání takto zpřístupněných prací přejímá repozitář od vydavatele.

Na webu repozitáře nebyl nalezen závazek k dlouhodobé archivaci uloženého obsahu, ani provedení auditu fungování repozitáře.

Na příkladu tohoto institucionálního repozitáře je vidět, jak důležitého partnera představuje management univerzity. Ústřední knihovna VŠB-TUO je asi nejangažovanější akademickou knihovnou v České republice v otázce otevřeného publikování. Technická implementace repozitáře je na velmi vysoké úrovni. Přesto není repozitář zakotven v žádné ze směrnic a získávání informací o vědecké činnosti probíhá bokem z vlastních zdrojů Ústřední knihovny.

4.11 Repozitář publikační činnosti Univerzity Tomáše Bati ve Zlíně

UTB se také rozhodla postavit svůj institucionální repozitář na software DSpace. Repozitář je spravován Knihovnou UTB⁸⁹ a je specializován pouze na publikační činnost univerzity. Mandát repozitáře pro archivaci výstupu autorů na univerzitě je řešen Rozhodnutím rektora č. 10/2015⁹⁰, kde se v odstavci 8 čl. 2 píše:

„Při doplňování údajů u jednotlivých záznamů v OBD je autor povinen jako přílohu přiložit postprint příslušného článku. Ten bude následně uložen v Repozitáři publikační činnosti UTB a zpřístupněn podle statusu časopisu v databázi Sherpa/Romeo. Status daného časopisu prověří Knihovna UTB. V případě, že vydavatel neumožňuje volné zveřejnění postprintu článku, bude tento článek v repozitáři dostupný pouze v rámci počítačové sítě UTB.“

U záznamů v repozitáři lze díky napojení na citační databáze a altmetriky zjistit zajímavé informace o využívání daného dokumentu ve vědecké komunitě. Nicméně není u záznamů specifikováno, pod jakou licencí jsou zpřístupňovány veřejnosti a to ani u záznamů v sekci Open Access. Repozitář nemá stanoven závazek k dlouhodobé archivaci, archivační politiku ani provedení auditu.

⁸⁹ Knihovna Univerzity Tomáše Bati ve Zlíně: <http://www.utb.cz/knihovna>

⁹⁰ Rozhodnutí rektora č. 10/2015: <http://www.utb.cz/file/51348/>

Univerzita provozuje separátně institucionální repozitář Digitální knihovna UTB⁹¹ určený na kvalifikační práce.

4.12 Národní úložiště šedé literatury

Repozitář NUŠL vznikl na základě projektu Digitální knihovna pro šedou literaturu - funkční model a pilotní realizace⁹² řešeného Národní technickou knihovnou a Vysokou školou ekonomickou v Praze. V rámci projektu byla vydána publikace Repozitáře šedé literatury, která shrnuje tehdejší situaci repozitářů pro šedou literaturu v České republice, legislativu spojenou s tímto druhem literatury, použitá metadata i popis tehdy aktuálních softwarů pro implementaci. Na knize se podílelo mnoho autorů napříč českými i zahraničními univerzitami. (Vaska et al., 2010) Druhým důležitým dokumentem, který z projektu vzešel je Metodika zpracování, dlouhodobého uchování a zpřístupnění šedé literatury v ČR na příkladu Národního úložiště šedé literatury. Tento document prakticky popisuje fázi definování repozitáře a jeho obsahu, výběru metadatového formátu a software a zajištění dlouhodobého uchování a zpřístupnění dat v repozitáři. Jak zmiňuje samotná autorka metodiky (Pejšová, 2011):

Cílem metodiky není vytvořit obecný návod na budování digitálního repozitáře, ale ukázat praktické postupy při vytváření digitálního úložiště pro šedou literaturu a vymezit jeho nastavení způsobem, který odpovídá podmínkám pro připojení do národních a mezinárodních systémů šedé literatury.

Tento repozitář provozuje Národní technická knihovna a je postaven na softwaru Invenio⁹³. Nicméně vyhledávání a indexování sítě repozitářů zapojených do projektu je zajištěno separátním systémem Elasticsearch⁹⁴. (NUŠL, 2017)

V repozitáři se nachází velká škála typů dokumentů⁹⁵, ve zkratce jsou to: kvalifikační práce, zprávy, autorské práce, analytické a metodické materiály, propagační a vzdělávací materiály, konferenční materiály a firemní literatura. Repozitář má nastavenou a zveřejněnou politiku pro příjem obsahu. Dále vyjadřuje závazek dlouhodobé archivace a nabízí jej jako benefit pro své

⁹¹ Digitální knihovna UTB: <http://digilib.k.utb.cz>

⁹² <http://repozitar.techlib.cz/record/407>

⁹³ Digitální repozitář NUŠL: <http://invenio.nusl.cz/>

⁹⁴ Národní úložiště šedé literatury: <http://www.nusl.cz/>

⁹⁵ Podrobnější informace na: <https://nusl.techlib.cz/nusl/typologie-dokumentu-nusl/>

uživatele, což dokládá absolvováním interního auditu nástrojem DRAMBORA⁹⁶. Tento audit byl však naposledy proveden v roce 2012 a dle dostupných informací nebyl od té doby proveden žádný jiný audit.

4.13 Studijní knihovna pro výzkum vývoje NTK

Tento repozitář je také postaven na systému Invenio a je spravován zaměstnanci Národní technické knihovny. Obsah repozitáře je tvořen šedou literaturou, jejímiž autory a autorkami jsou zaměstnanci NTK. Povinnost zaměstnanců ukládat své práce do repozitáře stanovuje Směrnice č. 1/2011⁹⁷. Ve směrnici je také stanovena licence, pod kterou je práce v repozitáři zveřejňována a to Creative Commons BY-NC-ND, případně BY-NC-SA. Repozitář nezmiňuje závazek k dlouhodobé archivaci svého obsahu, ani aplikaci auditu na svůj provoz. (Národní technická knihovna, 2017a)

4.14 Repozitář Akademie věd České republiky

O provoz tohoto repozitáře se stará Knihovna Akademie věd ČR⁹⁸. Repozitář je rozšířením databáze ASEP, která obsahuje záznamy o výsledcích základního výzkumu v Akademii věd ČR, a byl zprovozněn v roce 2013. Repozitář je tedy postaven na technologii ARL od firmy Cosmotron s.r.o. Dle Směrnice č. 4/2013⁹⁹ je povinné přidávat do databáze ASEP veškerý vědecký výstup AV ČR, nicméně ukládání plných textů je dobrovolné. Přidávání plných textů do repozitáře se řídí metodikou ukládání plných textů a je v praxi prováděno proškolenými osobami, tzv. zpracovateli. Od roku 2017 je možné do repozitáře nahrávat také datasety. Každý ústav, který chce do repozitáře ukládat plné texty musí mít podepsanou rámcovou licenční smlouvu. Při ukládání je uživateli nabídnuto zpřístupnit plný text volně, volně s embargem, pro ústav zaměstnance, nebo jen pro nahrávajícího zaměstnance. Pokud je plný text zveřejněn, jeho užití podléhá autorskému zákonu a je tedy možné jej užít pouze pro osobní potřebu. (Knihovna Akademie věd ČR, 2017a) Licence Creative Commons je možno vybrat pouze u uložených datasetů. (Knihovna Akademie věd ČR, 2017b)

⁹⁶ Audit důvěryhodnosti Digitálního repozitáře NUŠL: <https://nusi.techlib.cz/nusi/audit-drambora/>

⁹⁷ Směrnice č. 1/2011 ředitele NTK o publikační činnosti zaměstnanců NTK a právech k výsledkům VaVaI: <http://repozitar.techlib.cz/record/85?ln=cs>

⁹⁸ Knihovna Akademie věd ČR: <https://www.lib.cas.cz/>

⁹⁹ Směrnice Akademické rady AV ČR – Tvorba a využívání databáze ASEP: https://www.lib.cas.cz/asep/smernice/smernice_4_2013.pdf

Na webu repozitáře, Knihovny AV ČR, ani dodavatelské firmy¹⁰⁰ nebyl nalezen závazek k dlouhodobé archivaci obsahu ani provedení auditů.

4.15 Digitální knihovna Západočeské univerzity v Plzni

Tento institucionální repozitář obsahuje jak publikační činnost akademických pracovníků, tak kvalifikační práce. Tyto základní dva typy obsahu jsou uloženy každý v jiné „komunitě“ a lze mezi nimi snadno navigovat. Obsah se dále dělí do „podkomunit“ podle fakult, a kolekcí podle typu dokumentu. Repozitář je postaven na softwaru DSpace ve verzi 5.4, který poskytuje přívětivé grafické rozhraní a možnost filtrování obsahu podle autora, předmětu, data vydání či typu dokumentu.

Zveřejňování kvalifikačních prací v repozitáři se řídí Směrnicí rektora č. 26R/2013. Bohužel tato směrnice není veřejně dostupná a nelze tedy dohledat konkrétní podmínky. Ukládání publikační činnosti se řídí Směrnicí rektora č. 21R/2011¹⁰¹, která zmiňuje pouze ukládání papírové verze výstupu.

Od listopadu 2016 do března 2017 probíhal pilotní provoz napojení repozitáře na systém OBD s podporou automatizovaného nahrávání plných textů do repozitáře. Autor při vyplňování povinných polí nahrává také plný text práce (preprint, postprint, publikovaná verze) a systém podle dat ze služby SHERPA/ROMEO automaticky načítá příslušnou politiku daného odborného časopisu. (Univerzitní knihovna ZČU, 2016) Nicméně stav v době psaní této práce je nejasný. V repozitáři jsou sice vidět záznamy publikační činnosti s plnými texty, nicméně není jasné podle jaké metodiky a jak velká část univerzity se na naplňování repozitáře podílí.

Licenční práva pro využívání obsahu repozitáře se řídí autorským zákonem České republiky, a zároveň licenčními podmínkami časopisů, kde byly publikovány. Repozitář nevyužívá licencí Creative Commons a nezveřejňuje závazek k dlouhodobé archivaci ani prodělení auditu.

¹⁰⁰ ARL - CONTENTSERVER - PŘIPOJENÍ DIGITÁLNÍCH OBJEKTŮ:
<http://www.cosmotron.cz/contentserver-pripojeni-digitalnich-objektu>

¹⁰¹ Směrnice rektora č. 21R/2011 - Evidence publikačních činností a dalších odborných aktivit:
<http://www.knihovna.zcu.cz/export/sites/knihovna/publikace/dokumenty-publikacni-cinnost/smernice-o-publikacni-cinnosti.pdf>

5 Současný stav na České zemědělské univerzitě v Praze

5.1 Institucionální repozitář na České zemědělské univerzitě

Ze zákona je každá vysoká škola povinna zveřejňovat kvalifikační práce v tzv. databázi kvalifikačních prací. Nicméně provedení samotného zveřejňování, ať už v repozitáři či jiném online řešení, je ponechána značná svoboda. Tuto svobodu zajišťuje sám vysokoškolský zákon, kde se v § 47b od. 1 zákona č. 111/1998 Sb., o vysokých školách píše: „*Vysoká škola nevýdělečně zveřejňuje bakalářské, diplomové, disertační a rigorózní práce, u kterých proběhla obhajoba, včetně posudků oponentů a záznamu o průběhu a výsledku obhajoby prostřednictvím databáze kvalifikačních prací, kterou spravuje. Způsob zveřejnění stanoví vnitřní předpis vysoké školy.[...]*” (Česká republika, 1998) Proto se můžeme v praxi setkat s tím, že jedna vysoká škola zveřejňuje všechny kvalifikační práce, posudky oponentů a záznam o průběhu obhajoby bez jakýchkoli zábran a registrací návštěvníků, zatímco jiná zveřejní pouze metadata o pracích, ale pro přístup k samotným pracím a zmíněným dokumentům je potřeba registrace, kontaktování Help Desku nebo dokonce osobní návštěva knihovny.

Podobná situace je také na České zemědělské univerzitě. Repozitář je zde řešen v rámci modulu Univerzitního informačního systému¹⁰². Toto řešení umožňuje snadné propojení procesu nahrání dat s následným zveřejněním po obhajobě, protože vše se děje v rámci jednoho systému. Repozitář také umí exportovat záznamy a je sklizen Národním úložištěm šedé literatury¹⁰³, nicméně repozitář nemá proaktivní přístup ani ambici být sklizen dalšími agregátory dat. Samotné zveřejňování kvalifikačních prací popisuje čl. 9 Směrnice rektora č. 5/2017¹⁰⁴:

(2) Veřejně je na portálu veřejných informací UIS přístupný: typ práce, název práce, autor, vedoucí práce, garantující pracoviště, oponent, datum obhajoby, jazyk práce, abstrakt práce a klíčová slova. Tyto údaje jsou dostupné v jazyce závěrečné práce a v anglickém jazyce (pokud je práce zpracována v anglickém jazyce, tak jsou dostupné pouze v anglickém jazyce). Dle těchto parametrů je možné na této webové stránce závěrečné práce třídit a vyhledávat.

¹⁰² Univerzitní informační systém – Závěrečné práce: <http://is.czu.cz/zp/>

¹⁰³ Národní úložiště šedé literatury: <http://nysl.cz/?offset=0&zdroj=Školství%2FVysoké+školy%2FČeská+zemědělská+univerzita+v+Praze>

¹⁰⁴ Směrnice rektora č. 5/2017 - Pravidla zadávání, zpracování, odevzdávání, archivace a zveřejňování bakalářských a diplomových prací na ČZU: <https://www.czu.cz/cs/r-7210-o-czu/r-7702-oficialni-dokumenty/r-7810-vnitri-predpisy-univerzity>

(3) Pro přihlášené uživatele UIS, je dostupná i plná verze závěrečné práce ve formátu pdf, zadání, posudek vedoucího, jméno a posudek oponenta a výsledek obhajoby. Výjimkou jsou informace, kde je odloženo zveřejnění dle čl. 10 této směrnice.

(4) Pro zájemce mimo ČZU je plná verze závěrečné práce a další dokumenty podle odst. 3 tohoto článku zpřístupněna na základě žádosti zaslané na Helpdesk ČZU.

Návštěvník, který by se chtěl podívat na plné texty, musí tedy mít v systému vytvořen účet. Z toho vyplývá, že pokud není studentem ani zaměstnancem, musí se ohlásit na Helpdesku ČZU a o přístup si zažádat. Bez přihlášení do systému jsou totiž vidět pouze bibliografická data. Navíc v praxi je toto žádání o přístup implementováno úplně jinak. Na stránkách repozitáře po vyhledání kvalifikační práce je uživatel pro přístup k neveřejným souborům odkázán na vedoucí konkrétní práce. Není tedy jasné, zda vedoucí má za úkol poskytnout komukoli tyto neveřejné soubory na vyžádání nebo dát požadavek na Helpdesk ČZU, aby byly vygenerovány přístupové údaje. V prvním případě postup neodpovídá pokynům popsaným ve výše zmíněné Směrnici rektora č. 5/2017. Ve druhém případě je to jakási oklika, kdy se zbytečně zatěžuje vedoucí práce kvůli systémovému nastavení.

Na univerzitě není v současné době v provozu žádný institucionální repozitář pro vědce, kde by mohli archivovat a zpřístupňovat své preprinty, postprinty či práce samotné, pokud by to licence vydavatele povolila.

5.2 Dotazníkové šetření mezi zaměstnanci a doktorandy ČZU

Cílem tohoto dotazníkového šetření je zmapovat chování doktorandů a vědců na České zemědělské univerzitě za účelem porozumění jejich vztahu k otevřenému přístupu a vlastnímu ukládání vytvořených prací.

Respondenti odpovídají na dva typy otázek. Jeden typ jsou otázky, kde vybírají z předpřipravených odpovědí tu, která nejvíce vyhovuje. Tento typ zkoumá faktické otázky v první části jako věk, obor, frekvenci publikování atp. Druhým typem je otázka s odpovědí na Likertově škále. Respondenti mají na výběr škálu z pěti bodů – od rozhodně nesouhlasím, až po rozhodně souhlasím. Dotazník je složen ze tří částí a je přístupný v příloze této práce.

První část tvoří série sedmi otázek a zkoumá faktické postavení respondentů – jejich věk, obor výzkumu, akademickou hodnost, frekvenci publikování a zkušenosti s ukládáním do nějakého institucionálního repozitáře. Statisticky by se tyto proměnné daly popsat jako nominální a ordinální. Jedna z úvodních otázek se zabývá výzkumným oborem respondenta. Aby byla zajištěna větší šance na shodu v oboru více respondentů je zde nabídnut výběr z číselníku

Ministerstva školství, mládeže a tělovýchovy podle ISCED-97. Nejprve byly vybrány hlavní skupiny oborů¹⁰⁵, které se vztahují k ČZU. Podle těchto byly následně vybrány podskupiny oborů¹⁰⁶. Respondent má možnost zvolit „Jiné“, pokud není jeho obor ve výčtu uveden.

Druhá část je založena na výzkumu Kimové (2010) a zkoumá motivace a individuální nebo oborové bariéry pro svou vlastní archivaci vědci v režimu otevřeného přístupu. Proměnné v této části jsou ordinální na pětibodové škále „zcela souhlasím“ až „zcela nesouhlasím“. Pro potřeby tohoto výzkumu jsou hodnotám na škále přiřazeny celočíselné hodnoty 1 až 5. Z odpovědí na otázky z druhé části bude sestaveno výsledné „skóre k otevřenému přístupu“, které pak bude využito pro porovnání skupin respondentů podle proměnných v první části.

Poslední částí dotazníku je opět sled otázek s Likertovou škálou, které zjišťují, co by vědce motivovalo k tomu, aby do institucionálního repozitáře přispívali. Tyto otázky vychází z následného výzkumu Kimové (2011).

5.3 Výběr otázek

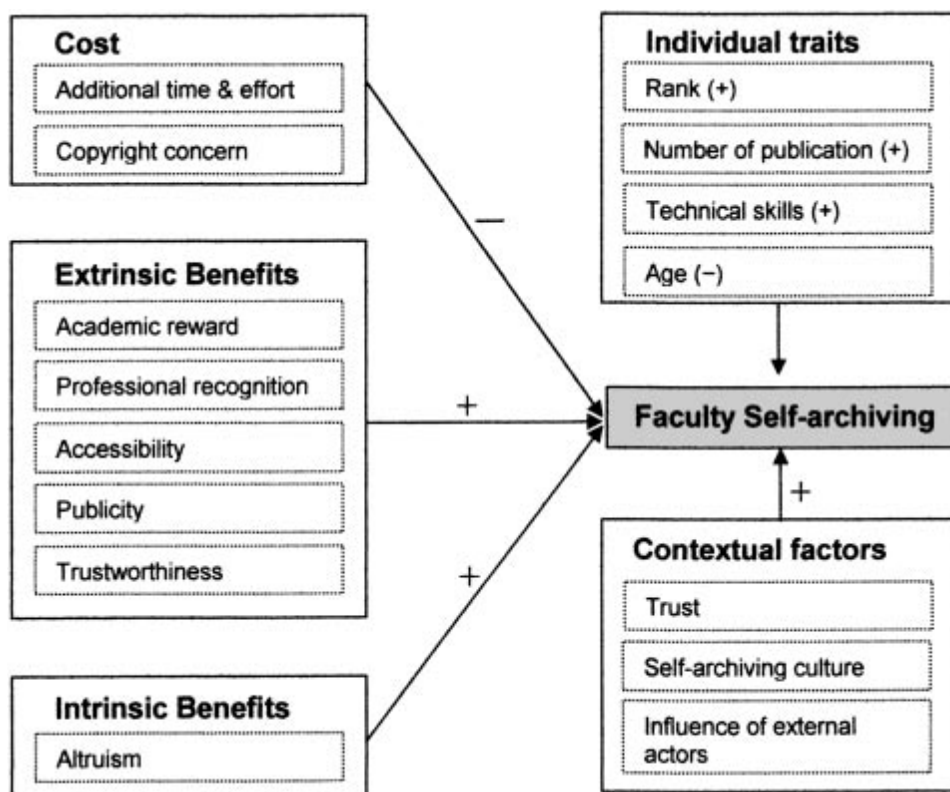
Kimová ve svém výzkumu přístupu a vnímání vlastní archivace vědců v akademické sféře zkoumala vliv celkem patnácti faktorů na chování v rámci vlastní archivace. Tyto faktory nejsou vybrány náhodně, nýbrž na základě rozsáhlé studie zdrojů zabývajících se tímto tématem. Na Ilustraci 5: Diagram faktorů a jejich předpokládaného pozitivního či negativního vlivu na chování self-archivingu. (Kim 2010) je znázorněn původní model, který Kimová vytvořila na základě existující literatury.

Na základě korelace dat z dotazníku s reálným chováním vědců poté byla schopna stanovit nejvlivnější faktory na chování při vlastní archivaci a to na celkem osm – altruismus, kultura vlastní archivace v oboru, obavy z porušení autorského práva, technické dovednosti jedince, věk, akademický zisk, obava z potřeby dodatečného času a námahy a obor respondenta. Z jejího výzkumu také vyplývá, že kultura vlastní archivace a technické dovednosti jsou podmíněné konkrétním oborem respondenta. Na Ilustraci 6 je vidět výsledný model těchto faktorů majících

¹⁰⁵ AKF1 Hlavní skupiny oborů ISCED-97 (jednomístné kódy) - <http://stistko.uiv.cz/katalog/cslnk.asp?idc=AKF1&ciselnik=Hlavn%ED+skupiny+obor%F9+ISCED-97+%28jednom%EDstn%E9+k%F3dy%29&aap=on>

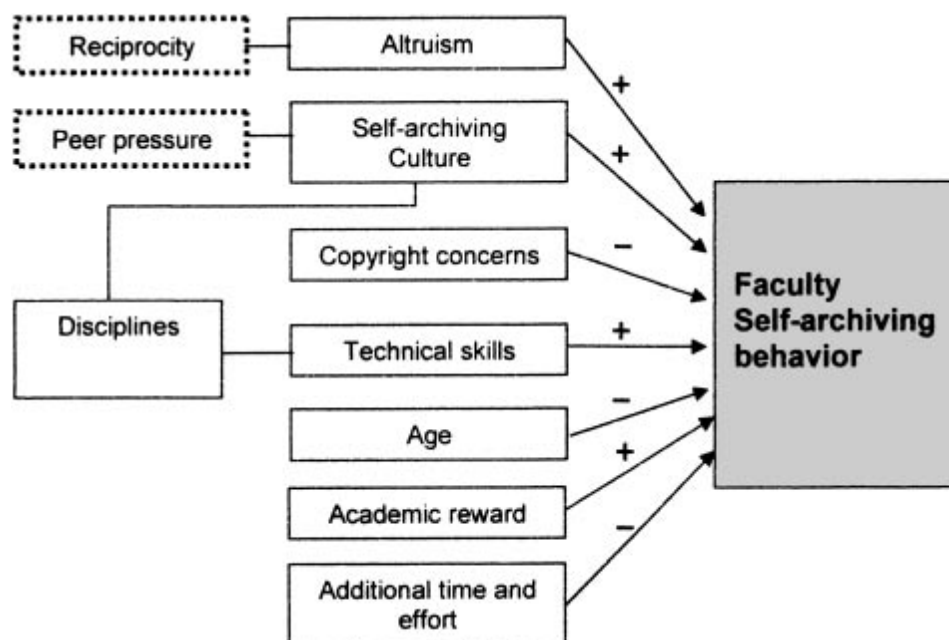
¹⁰⁶ AKF3 Podskupiny oborů ISCED-97 (trojmístné kódy) - <http://stistko.uiv.cz/katalog/cslnk.asp?idc=AKF3&ciselnik=Podskupiny+obor%F9+ISCED-97+%28trojm%EDstn%E9+k%F3dy%29&aap=on>

kladný nebo negativní vliv na vlastní archivaci podle Kimové. Na základě této studie jsou v průzkumu v praktické vybrány otázky zkoumající těchto osm oblastí.



Ilustrace 5: Diagram faktorů a jejich předpokládaného pozitivního či negativního vlivu na chování self-archivingu. (Kim 2010)

Kimová (2011) dále rozšířila svou práci i na výzkum motivace respondentů pro ukládání vědeckých prací přímo do institucionálního repozitáře. Tedy, co by je přimělo k tomu, aby svou práci archivovali či vystavili právě v institucionálním repozitáři. Proto jsou v tomto dotazníkovém šetření zařazeny i některé otázky z tohoto výzkumu. Díky konkrétnosti těchto otázek lze poté při implementaci repozitáře lépe stanovit požadavky na výslednou funkcionalitu systému podle požadavků akademické komunity instituce.



Ilustrace 6: Výsledný model hlavních faktorů majících vliv na self-archiving (Kim 2010)

5.4 Metodologie

Výzkumné otázky dotazníkového šetření jsou z pohledu statistiky tvořeny nominálními a ordinálními proměnnými. Nominální proměnné neposkytují prostor pro mnoho matematických operací a v tomto výzkumu budou využity pro nahlížení na soubor respondentů z různých perspektiv. Jsou to otázky zjišťující věk, akademickou hodnost, obor, úroveň práce s počítačem, publikační činnost a zkušenosti s publikováním mimo odborný časopis.

Ordinální proměnné v tomto výzkumu jsou tvořeny výhradně otázkami s odpověďmi na Likertově škále a nachází se ve druhé a třetí části dotazníku. Ve druhé části je s jejich pomocí zkoumán vztah respondenta k vlastnímu archivování a otevřenému přístupu k vědě v tomto kontextu. Článek Carifia a Perly (2008) shrnuje vývoj přístupu ke statistickému vyhodnocení otázek s Likertovou škálou za posledních 50 let. Zpracování tohoto statistického nástroje totiž obecně budí určitou kontroverzi, neboť v určitých případech lze chápat proměnné s Likertovou škálou jako intervalovou proměnnou. Právě podle jejich výkladu a mnoha dalších jimi zmíněných studií lze tedy skupinu více otázek Likertovy škály na podobné téma chápat jako intervalové proměnné.

V návaznosti na článek Carifia a Perly (2008) a protože výzkum Kimové již vybral otázky, jejichž zodpovězení pozitivně nebo negativně koreluje s postojem k open-accessu a vlastnímu archivování, lze si dovolit chápat odpovědi na otázky ve druhé části dotazníku jako intervalové. To dává prostor k provedení široké škály matematických operací nad těmito proměnnými, například určit rozdíly mezi obory.

Proměnné ve třetí části zkoumají motivaci k ukládání do institucionálního repozitáře a budou vyhodnoceny jako ordinální proměnné.

5.4.1 Získání dat a stanovení hypotéz

V praktické části práce byl proveden kvantitativní výzkum mezi vědci na České zemědělské univerzitě. Bohužel vzhledem k tomu, že návratovost dotazníků byla tak nízká, nepodařilo se získat dostatek respondentů pro provedení kvalitní a věrohodné kvantitativní analýzy.

Nicméně i přes to se autor práce rozhodl postup vyhodnocení výsledků zahrnout do této práce pro případ, že by se někdo další rozhodl v budoucnu výzkum replikovat. Případně pokud by byla vůle tento výzkum zopakovat na ČZU.

Tento průzkum si klade za cíl jednak získat podklady pro sepsání doporučení pro vedení univerzity pro budování institucionálního repozitáře. Dalším cílem je ověření jednoho z výsledků Kimové, že obor respondenta má vliv na jeho příklon či odklon k myšlence otevřeného přístupu. Pro tyto účely jsou stanoveny dvě hypotézy:

H₀: Neexistuje závislost mezi oborem respondenta a nadprůměrností či podprůměrností OA skóre.

H₁: Existuje vztah mezi oborem respondenta a nadprůměrností jeho OA skóre.

Zamítnutí či potvrzení nulové hypotézy H₀ bude provedeno za pomoci chí-testu. Tedy vypočtením chí-kvadrátu pro získané obory a jeho porovnáním s kritickou hodnotou pro hladinu významnosti 0,05 získanou z tabulek kritických hodnot pro chí-kvadrát. Aby bylo možno aplikovat chí-test, byla zavedena nová intervalová proměnná OA_skore, která je vypočtena jako suma kódovaných odpovědí na otázky 1 až 15 Likertovy škály v dotazníku. Na základě vypočtených OA_skore pro respondenty je vypočteno průměrné OA_skore pro celý vzorek jako:

$$prum_OA_skore = \frac{\sum_{i=1}^n OA_skore_i}{n}$$

Kde n ... počet respondentů; i ... index řádku respondenta

Dále je zavedena nominální proměnná nadprum_OA_skore, která určuje, zda OA_skore daného respondenta je vyšší než prum_OA_skore, v tom případě bude mít hodnotu PRAVDA, nebo zda je stejná či menší. V tom případě bude hodnota proměnné NEPRAVDA.

Výsledná tabulka bude poté vypadat podle následujícího schématu:

	Nadprůměrný OA				
Obor	P_NEPRAVDA	O_NEPRAVDA	P_PRAVDA	O_PRAVDA	Celkem
Obor i	$P_{NEPRAVDA_i}$	$O_{NEPRAVDA_i}$	P_{PRAVDA_i}	O_{PRAVDA_i}	
Celkem	$\sum_{i=1}^n P_{NEPRAVDA_i}$		$\sum_{i=1}^n P_{PRAVDA_i}$		$\sum_{i=1}^n P_{NEPRAVDA_i} + \sum_{i=1}^n P_{PRAVDA_i}$

Tabulka 3: Schéma výpočtu pozorovaných četností a očekávaných četností pro jednotlivé obory

Kde P ... pozorované četnosti; O ... očekávané četnosti

Způsob výpočtu očekávaných četností O je následující:

$$O_{NEPRAVDA_i} = \frac{\sum_{i=1}^n P_{NEPRAVDA_i}}{\sum_{i=1}^n P_{NEPRAVDA_i} + \sum_{i=1}^n P_{PRAVDA_i}} \times (P_{NEPRAVDA_i} + P_{PRAVDA_i}),$$

kde i ... index řádku oboru; n ... počet všech oborů

5.4.2 Úprava dat

Po skončení sběru dat se může stát, že nebude dostatek respondentů u vyplněných oborů. Pokud se nezapojí do výzkumu odhadem více než stovka, spíše několik stovek respondentů, je tento scénář velice reálný. V takovém případě bude přikročeno k překódování oborů všech respondentů podle nadřazené úrovně vybraných oborů číselníku MŠMT¹⁰⁷. Tímto krokem se zredukuje počet řádků při výpočtu chí-kvadrátu a zvýší počet pozorovaných četností. Bude tedy možné použít nižší kritickou proměnnou pro porovnání s chí-kvadrátem.

5.4.3 Propagace dotazníku a výběr vzorku

Dle výroční zprávy za rok 2016 studuje na univerzitě 797 doktorandů v prezenčním i dálkovém studiu a pracuje zde 788 akademických a vědeckých pracovníků. (Česká zemědělská univerzita, 2017) Dohromady tedy 1 585 potenciálních respondentů. K této populaci se však nepodařilo získat přístup. Navíc v souvislosti s blížícími se volbami nového rektora ČZU se rozhodlo vedení univerzity co nejvíce omezit hromadné rozesílání emailů, které by mohly akademickou obec a zejména pedagogy a vědecké pracovníky obtěžovat.

Při konzultaci s prorektorkou pro vědu a výzkum prof. Mgr. Ing. Markétou Sedmíkovou, Ph.D. o možnosti kontaktování akademických pracovníků a vědců na České zemědělské

¹⁰⁷ AKF1 Hlavní skupiny oborů ISCED-97 (jednomístné kódy) - <http://stistko.uiv.cz/katalog/cslnk.asp?idc=AKF1&ciselnik=Hlavn%ED+skupiny+obor%F9+ISCED-97+%28jednom%EDstn%E9+k%F3dy%29&aap=on>

univerzitě byla nabídnuta možnost hromadně obeslat všechny řešitele projektů CIGA¹⁰⁸. Dle 7. odstavce 2. článku Rozhodnutí rektora č. 2/2017¹⁰⁹ jsou řešitelé projektů doktorandi, akademičtí pracovníci ČZU nebo postdoktorandi. Výsledný vzorek výzkumu je tedy podmnožinou populace 1585 osob, nicméně nepodařilo se přesně zjistit, jak velký tento vzorek je. Zároveň v textu e-mailu muselo být zdůrazněno, že je vyplnění dotazníku nepovinné.

Kritickým bodem každého dotazníkového šetření je způsob oslovení zkoumaného vzorku a komunikační kanál. Důležitá je formulace oslovení tak, aby potenciální respondenti získali o vyplnění dotazníku zájem nebo alespoň, aby zájem o vyplnění hned neztratili. Neméně důležitý je výběr správného kanálu komunikace. Kanálů se nabízí několik: osobní návštěva, dopis, telefonát či e-mail. U každého z těchto typů komunikačních kanálů je pak nutné uvážit, zda je efektivní vztah vynaložených zdrojů pro oslovení zkoumaného vzorku vůči účinnosti daného kanálu. Například osobní návštěva je účinná, protože navštívená osoba poté nejspíše dotazník vyplní, avšak zároveň je to metoda velmi náročná na zdroje.

Pro účely tohoto dotazníkového šetření bylo sáhnuto po kanálu e-mail s následným vyplněním samotného dotazníku na platformě Google Forms¹¹⁰. Oslovovaný vzorek je příliš velký, než aby bylo možné zvládnout v reálném čase osobní návštěvy či rozesílání dopisů. Pro tuto metodu komunikace mluví také fakt, že každý zaměstnanec univerzity má přidělený e-mail.

¹⁰⁸ Celouniverzitní interní grantová agentura ČZU v Praze

¹⁰⁹ Vyhlášení soutěže Celouniverzitní interní grantové agentury ČZU v Praze (CIGA) pro rok 2018 – Rozhodnutí rektora č. 2/2017:

<https://ga.czu.cz/NLDefault/NewsArticle.aspx?id=2361995&backTo=https%3a%2f%2fga.czu.cz%2fdefault.aspx%3f%26page%3d1%26page%3d1%26page%3d1>

¹¹⁰ Formuláře Google: https://www.google.com/intl/cs_ALL/forms/about/

6 Návrh postupu pro tvorbu institucionálního repozitáře na ČZU

6.1 Výsledky průzkumu na ČZU

Průzkumu se celkem zúčastnilo 26 respondentů. Nejedná se tedy ani zdaleka o reprezentativní vzorek, který by zastoupil celou akademickou sféru univerzity. Proto je třeba při čtení výsledků průzkumu vzít tento fakt v úvahu, protože jej nelze jednoznačně zobecnit na celou populaci akademických pracovníků České zemědělské univerzity zahrnující doktorandy a akademické a vědecké pracovníky.

Důvodem pro tak nízkou návratovosti může být styl, kterým musel být text e-mailu napsán, a také nemožnost přímo oslovit všechny akademické pracovníky univerzity, nýbrž pouze podskupinu v registru žadatelů o univerzitní granty. Bohužel v důsledku malé návratnosti dotazníků se také nepodařilo vyvrátit hypotézu H_0 , a tedy ani potvrdit, zda obor respondenta má vliv na jeho postoj k otevřenému přístupu. Výpočet chí-kvadrátu lze nalézt v příloze této práce.

Dle vyplněných dat tvoří 73 % respondentů osoby s akademickou hodností magistr/inženýr, autor práce předpokládá, že se jedná o doktorandy. Zbýlých 27 % má akademickou hodnost doktor. Vzhledem k problematice otevřeného přístupu je zajímavá odpověď na otázku *Už jsem někdy zpřístupnil plný text své práce veřejně skrze nějakou jinou instituci než vydavatele časopisu?*, která zkoumá předchozí zkušenost respondentů s otevřeným přístupem ke své práci. Z odpovědí vyplývá, že 58 % respondentů má zkušenost s publikováním v režimu otevřeného přístupu, ať už v oborovém repozitáři, na osobním nebo fakultním webu, či na sociálních sítích jako ResearchGate či Mendeley.

V Tabulka 4 jsou shrnuty odpovědi na otázky zkoumající motivaci respondentů k ukládání do institucionálního repozitáře. Vedle celkového počtu respondentů s danou mírou souhlasu či nesouhlasu je pro lepší představu vypsáno i procentuální vyjádření z celkového počtu 26 respondentů. K problematice zveřejňování preprintů prací se vyjadřuje komentář jednoho z respondentů:

„Zpřístupňování rukopisů vidím jako velmi problematické (1) kvůli možnému ukradení výsledků, (2) na článku se podílí také recenzenti, kteří jej často pomohou značně vylepšit. Rukopis a následnou publikaci nelze zcela srovnávat.“

Z odpovědí lze vyčíst, že respondenti mají zájem o statistiku zobrazení a stažení jejich prací, pro což se alespoň souhlasně vyjádřilo 88 % respondentů. Respondenty by dále motivovalo formální uznání přínosu vědců ukládajících do institucionálního repozitáře ze strany univerzity, alespoň částečný souhlas vyjádřilo 84 % respondentů. 73 % respondentů by alespoň částečně motivovalo, pokud by ukládání do institucionálního repozitáře bylo počítáno do habilitačního a jmenovacího řízení či doktorského studia. Na konto této problematiky se vyjádřil jeden z respondentů ve svém komentáři následovně:

„Některé z uvedených otázek jsou v přímém rozporu s otevřeným přístupem např. dotaz na ocenění (na úrovni ČZU) za zpřístupnění výstupů. Oceňovat by měli Ti, kteří článek využijí, nikoliv instituce. Zároveň se domnívám, že uznávání výstupů v otevřeném přístupu v rámci habilitačního a profesorského řízení je v rozporu s vysokoškolským zákonem.“

Otázka	Zcela nesouhlas	Částečně nesouhlas	Nevím	Částečně souhlas	Zcela souhlas
16. Kdyby dokázal spolehlivě uchovat uložené soubory po mnoho let. (15 let a více)	1 (4 %)	6 (23 %)	6 (23 %)	9 (35 %)	4 (15 %)
17. Kdyby bylo vidět, kolikrát si někdo moji práci zobrazil či stáhnul.	0	1 (4 %)	2 (8 %)	11 (42 %)	12 (46 %)
18. Kdyby nahrávání probíhalo automaticky nebo by se o to za mě staral někdo jiný.	2 (8 %)	1 (4 %)	4 (15 %)	7 (27 %)	12 (46 %)
19. Že by vydavatel nebyl jediný, kdo má kontrolu nad zpřístupněním mé práce.	2 (8 %)	0	9 (35 %)	8 (31 %)	7 (27 %)
20. Kdyby univerzita nějak formálně ocenila můj přínos do repozitáře.	1 (4 %)	1 (4 %)	2 (8 %)	5 (19 %)	17 (65 %)
21. Možnost ukázat můj výzkum, ještě než by byl dokončen, pro případnou kritiku od kolegů.	3 (12 %)	6 (23 %)	4 (15 %)	8 (31 %)	5 (19 %)
22. Kdyby se přispívání do IR počítalo do habilitačního	3 (12 %)	0	4 (15 %)	4 (15 %)	15 (58 %)

a jmenovacího řízení či doktorského studia.					
23. Kdybych dostal/a finanční odměnu.	1 (4 %)	1 (4 %)	6 (23 %)	3 (12 %)	15 (58 %)

Tabulka 4: Přehled odpovědí na otázky týkající se motivace ukládání prací do repozitáře. (n=26)

6.2 Stanovení klíčových skupin pro fungování repozitáře

Na fungování a nastavení repozitáře má vliv mnoho různých skupin osob. Autor práce navrhuje model, který z provedeného průzkumu politik institucionálních repozitářů na českých univerzitách převládá. Tedy o samotný provoz repozitáře se stará univerzitní či institucionální knihovna, v kontextu ČZU Studijní a informační centrum, a o funkčnost technického řešení se stará univerzitní IT odbor, v kontextu ČZU Odbor informačních a komunikačních technologií. Vzhledem k tomuto modelu lze identifikovat následující skupiny a jejich role:

- Management univerzity – tvorba mandátu, pověření a statusu repozitáře včetně souvisejících směrnic
- Odbor informačních a komunikačních technologií (dále OIKT) – zajištění provozu HW a SW; propojení s UIS; provádění zálohování a podílení se na řízení rizik v této oblasti
- Studijní a informační centrum (dále SIC) – nastavení procesů v repozitáři; kontrola metadat přijímaných objektů; revize použitých formátů; plány rozvoje; operační provoz repozitáře
- Akademičtí pracovníci – podílení se na tvorbě obsahu; zpětná vazba
- Vydavatelé – poskytnutí licence k publikovaným dokumentům a případně výzkumným datům
- Grantové agentury, MŠMT – stanovení obecných pravidel pro publikování veřejného výzkumu v repozitáři
- Veřejnost – užívání rozhraní repozitáře k získání informací a plných textů

6.3 Stanovení základních parametrů repozitáře

Repozitář, pokud má být obecně důvěryhodný pro své uživatele, potřebuje mít nastaveny základní parametry jako stanovení účelu a funkce, provozních cílů, velikosti a technického řešení. V této části budou za pomoci nástroje PLATTER popsány základní parametry, které by mohly být stavebním kamenem pro implementaci institucionálního repozitáře na ČZU.

6.3.1 Účel a funkce repozitáře

6.3.1.1 Mandát a pověření

Vzhledem k tomu, že repozitář by byl institucionální, je otázka mandátu celkem jasná. Mandát repozitáři by měla poskytnout sama instituce a jasně specifikovat, jak přispívá repozitář do její strategie výuky a výzkumu. Lze to vidět i na mnoha univerzitách v Česku, kde je institucionální repozitář zakotven ve směrnicích dané instituce.

6.3.1.2 Status

Otázka statusu řeší, zda by měl být repozitář ziskový či neziskový. Obecně je trendem, že institucionální repozitáře v případě univerzit jsou neziskovými. V otázce kvalifikačních prací toto vychází přímo ze zákona o vysokých školách. V otázce publikační činnosti by se v případě zpoplatnění jednalo buď o porušení licenčních smluv s vydavateli, nebo o provoz vlastního univerzitního nakladatelství.

6.3.1.3 Právní podmínky získávání obsahu

V kontextu vysokých škol je právní část získávání obsahu poměrně jednoduchá, protože v případě publikační činnosti se ukládají zaměstnanecká díla autorů a je tedy potřeba dát pozor pouze, pokud by se chtěla daná díla zveřejňovat. Zde je potřeba brát v úvahu i souhlas ostatních autorů, kteří nepůsobí na univerzitě, a také licenční podmínky primárních vydavatelů těchto děl. V otázce kvalifikačních prací studentů je tento problém opět komplexně řešen zákonem o vysokých školách.

6.3.1.4 Provozní vytrvalost

Repozitář na ČZU by se budoval od základu a tedy je možnost nastavit procesy a technologie tak, aby bylo už v návrhu počítáno s jeho důvěryhodností.

6.3.2 Velikost repozitáře

6.3.2.1 Množství dat

Otázka předpokládaného množství dat v repozitáři má zásadní vliv na použité technologie při jeho výstavbě a vynaložených zdrojích na jeho provoz a zálohování. V případě publikační činnosti by mohl výsledný počet nových prací ročně nastínit registr národní Informační systém výzkumu, experimentálního vývoje a inovací, konkrétně jeho rejstřík RIV¹¹¹. Do RIVu přibýlo za rok 2016 z České zemědělské univerzity 1 422 záznamů. Pokud by bylo

¹¹¹ REJSTŘÍK INFORMACÍ O VÝSLEDKÁCH: <https://www.rvvi.cz/riv>

možné do repozitáře nahrávat i neodbornou literaturu vznikající při výzkumu jako články v odborných periodických, různé reporty apod., mohlo by toto číslo ještě značně narůst. Dále Výroční zpráva o činnosti za rok 2016 uvádí, že ve zmíněném roce bylo celkem 5 088 úspěšných absolventů bakalářských, magisterských a doktorandských oborů. (Česká zemědělská univerzita, 2017) Národní úložiště šedé literatury, které sklízí i současný repozitář kvalifikačních prací na ČZU udává pro rok 2016 celkem 3 544 prací¹¹². Výsledný počet přírůstku kvalifikačních prací za rok 2016 bude pravděpodobně někde mezi těmito dvěma hodnotami. Pokud by tedy repozitář uchovával jak publikační činnost, tak kvalifikační práce, dohromady by za rok mohlo do repozitáře přibýt mezi 5 000 až 10 000 objekty.

6.3.2.2 *Lidské zdroje*

Otázka lidských zdrojů má dvě důležité části. První částí je samotné personální zajištění provozu repozitáře. Kolik úvazků bude k provozu potřeba a jaké know-how. Dobrou zprávou je, že provoz repozitáře není nepodobný práci knihovníka. V praxi by tedy mohlo stačit personální rozšíření stávajícího oddělení fondů knihovny SIC, které má na starosti také agendu katalogizace, a uspořádání školení pro práci se zvoleným software pro repozitář. Kolik bude potřeba přibrat úvazků, do velké míry závisí na způsobu napojení repozitáře na univerzitní informační systém. Budou muset zaměstnanci repozitáře pouze kontrolovat přijatá data nebo aktivně doplňovat další potřebná metadata? Například na Západočeské univerzitě v Plzni umí jejich informační systém automaticky doplnit i licenční politiku ohledně otevřeného přístupu časopisu, kde byl článek publikován. Dále je potřeba mít k dispozici alespoň na částečný úvazek osobu schopnou pracovat s licenčními politikami vydavatelů a v neposlední řadě také stálou IT podporu ve formě částečného úvazku.

Druhou částí otázky lidských zdrojů jsou samotní uživatelé. Je důležité mít představu, kolik unikátních uživatelů bude repozitář mít a přizpůsobit tomu infrastrukturu. Základním okruhem uživatelů by mohli tvořit absolventi, kteří minimálně využijí repozitář k uložení své práce, ale pravděpodobně také k prohlížení už existujících prací, pro rok 2016 tedy 5 088 osob. Tento okruh dále můžeme rozšířit o akademické pracovníky, kteří budou systém využívat jak k ukládání, tak k prohlížení práce svých kolegů a své vlastní, tedy 1 585 osob. (Česká zemědělská univerzita, 2017) Jen tyto dvě skupiny čítají dohromady 6 673 osob.

¹¹² Vyhledávací dotaz:

<http://nusi.cz/?language=cs&offset=0&zdroj=Školství%2FVysoké+školy%2FČeská+zemědělská+univerzita+v+Praze&from=2016&to=2016>

Vzhledem k tomu, že repozitář bude zahrnut v různých registrech jako DOAR či ROAR, indexovaný v Google Scholar atp., může se využívanost ještě znatelně zvýšit.

6.3.3 Provoz

6.3.3.1 Metody akvizice

Otázka získávání dat řeší, jak budou do repozitáře vstupovat objekty. V kontextu institucionálního repozitáře přichází v úvahu dva způsoby příjmu dat – externí systém a pracovníci repozitáře. Získávání dat z externího systému, např. univerzitní informační systém, se dle nástroje PLATTER nazývá metoda push. Tedy externí systém vkládá přes nějaké rozhraní data do systému. V tomto případě je na externím systému, aby zajistil kvalitu dat. Druhý způsob příjmu dat je zadávání pracovníky repozitáře, kteří pak ručí za správnost zadaných dat. Případně může jít o kombinaci těchto přístupů, kdy data jsou poslána z externího systému a poté ještě projdou kontrolou zaměstnanci repozitáře. V tomto případě je důležité, aby každý aktér v průběhu akvizice věděl, za která metadata zodpovídá, jako je to například specifikováno ve směrnici na Univerzitě Pardubice (viz podkapitola 4.9 Digitální knihovna Univerzity Pardubice).

6.3.3.2 Komplexnost dat

Základem pro dlouhodobou archivaci je dopředu vědět, jaké formáty dat bude repozitář přijímat. S touto informací je pak mnohem jednodušší naplánovat, jak se provede dlouhodobá archivace takových dat. Institucionální repozitář může očekávat soubory typu PDF, pokud se repozitář rozhodne pro archivaci výzkumných dat, může se seznam souborů rozšířit o obrázky, videa, tabulkové soubory, mapy, zdrojové kódy, exporty dat aj. Všechny tyto typy souborů mají ještě různé formáty, které se liší ve své vhodnosti k dlouhodobé archivaci¹¹³. V případě PDF souborů se prokázalo, že klasické PDF soubory nejsou vhodným formátem k dlouhodobé archivaci a proto byl vytvořen formát PDF/A¹¹⁴. Na repozitáři je nastavit pravidla pro příjem dat, například povinností uložitelů převést soubory do formátu PDF/A, jako na Karlově univerzitě, nebo případně automatické procesy pro převod uložených dat do vhodných formátů k dlouhodobé archivaci.

¹¹³ Jednoduchý přehled formátů: <https://documentation.library.ethz.ch/display/DD/File+formats+for+archiving>

¹¹⁴ Více na téma, proč je nutný standard PDF/A: <https://www.pdfa.org/pdfa-a-new-standard-for-long-term-archiving/>

6.3.3.3 Specializace dat

V případě institucionálního repozitáře a ukládání pouze PDF se nejedná o příliš specializovaná data. Nicméně i v případě například diplomových prací není důležité uchovat pouze práci samotnou, ale i některá klíčová metadata jako jméno školitele, oponenta, získaná známka, datum obhájení apod. V případě ukládání výzkumných dat je situace ještě složitější a rozmanitější, protože u každého typu souboru je nutné specifikovat, které jeho vlastnosti a metadata jsou důležité pro dlouhodobé uchování.

6.3.3.4 Citlivost dat

Repozitář může také obsahovat citlivá data komerční, lékařské či osobní povahy. Dle nástroje PLATTEER je nutné nastavit zabezpečení celého repozitáře podle nejcitlivějších dat uvnitř.

6.3.3.5 Oprávnění k přístupu

V kontextu institucionálního repozitáře bude nutné zajistit, aby bylo možné některé objekty před veřejností skrýt a to v různých režimech. Některé objekty budou veřejně dostupné bez omezení, některé jen pro zaměstnance a studenty univerzity, některé jen pro jejich uložitele. Do toho je nutné navíc zahrnout také možnost embarga od vydavatele, který může mít v licenčních podmínkách vyhrazeno zpoždění v řádu měsíců, po jehož uplynutí může být teprve soubor zveřejněn v repozitáři instituce.

6.3.4 Technická řešení a možnosti implementace

Na trhu existuje mnoho open source řešení pro repozitáře, které implementují mnohé ze zde zmíněných funkcí. Ať už je to interoperabilita, archivace zajišťující autenticitu a integritu dat aj. V prostředí České republiky převládají softwary DSpace a Invenio. Pro jejich funkčnost mluví také fakt, že většina českých univerzit provozuje své repozitáře právě na open source softwarech.

6.3.4.1 Zdroj metadat

Existuje několik možností, jak získat metadata popisující přijímaný objekt. V kontextu institucionálního repozitáře by hlavním zdrojem metadat měl být univerzitní informační systém. Je nutné si ale uvědomit, že i v případě získávání dat z informačního systému stojí na úplném začátku lidský faktor. Proto je nutné nastavit buď v informačním systému, nebo při příjmu dat do repozitáře příslušné kontroly vstupů a také kontrolu zadaných dat z externích zdrojů. Jako příklad lze uvést situaci, kdy se uživatel pouze vyplněním ISSN časopisu, kde publikoval svou práci, automaticky vyplní ostatní pole jako název časopisu, vydavatel a jeho

licenční politika. V případě zadávání metadat zaměstnanci repozitáře je nutné zajistit kontrolu kvality těchto dat a zajistit podmínky pro co nejmenší chybovost zaměstnanců.

6.3.4.2 *Standardy interoperability*

Velkou přidanou hodnotou institucionálního repozitáře je koncentrace výstupu instituce na jednom místě. Umožnění harvestování repozitáře externími subjekty jako Google Scholar, DOAR, ROAR apod. je logickým krokem k větší viditelnosti výstupu univerzity.

Institucionální repozitář by měl minimálně implementovat standard OAI-PMH sloužící právě ke sklizení dat z repozitáře. Dalšími standardy mohou být například SWORD¹¹⁵ pro strojové vkládání objektů do repozitáře či METS¹¹⁶ pro výměnu objektů mezi repozitáři.

6.3.4.3 *Strategie ukládání*

Ve fázi plánování nejsou až tak důležité technické podrobnosti ukládání, ale především to, kdo ponese za ukládání zodpovědnost. Má IT odbor dostatek know-how a zdrojů ke spolehlivé archivaci dat z repozitáře nebo bude správa dat svěřena externí firmě?

6.3.4.4 *Strategie softwarové podpory*

V kontextu České zemědělské univerzity mohou vyvstat tři možné strategie softwarové podpory. Pokud se univerzita rozhodne pro vývoj vlastního řešení, poskytne podporu pravděpodobně samotný IT odbor. Rozhodne-li se univerzita o provoz open source řešení, může být podpora získána opět z odboru IT, externí firmy specializující se na implementaci repozitáře nebo ze samotné komunity kolem vybraného software. Univerzita by se také mohla rozhodnout pro zakoupení komerčního software pro repozitář. Pak je samozřejmě nutné podporu zakotvit v příslušné smlouvě s dodavatelem.

6.4 Plány strategických cílů podle PLATTERu

Po vyjasnění základních otázek je možné přikročit k realizaci plánů plnících strategické cíle důvěryhodného repozitáře. Pro stanovení potřebných plánů bylo opět sáhnuto k nástroji PLATTER. V Tabulka 5 je zachycen seznam strategických plánů, které by měly být dle nástroje PLATTER vytvořeny. Tento seznam byl následně rozšířen o klíčové skupiny, které se budou na tvorbě těchto plánů podílet. Skupinou přítomnou ve všech případech je Studijní a

¹¹⁵ Více o standardu SWORD: <http://swordapp.org/about/>

¹¹⁶ Metadata Encoding & Transmission Standard: <https://www.loc.gov/standards/mets/METSOverview.html>

informační centrum, jakožto knihovna, která by se dle zkušeností z jiných univerzit měla starat o provoz repozitáře.

Plán strategických cílů	Odpovědnost	Klíčové skupiny
Finanční plán	Zabývá se plánováním, monitorováním a vykazováním financí.	SIC, management univerzity
Plán řízení lidských zdrojů	Zabývá se získáváním a udržováním souboru dovedností, které jsou relevantní pro správu repozitáře.	SIC, management univerzity
Datový plán	Specifikuje datové a metadatové objekty, formáty, struktury pro vkládání, uchovávání a zpřístupňování dat a související transformace a mapování dat.	SIC, OIKT, management univerzity
Akviziční plán	Zabývá se vztahy s depozitorem a dalšími poskytovateli dat.	SIC, OIKT
Plán zpřístupňování	Zabývá se vztahy s koncovými uživateli a pravidly pro zpřístupňování.	SIC, management univerzity, ak. pracovníci, nakladatelé, grantové agentury
Plán ochrany	Zajišťuje zpřístupňování a použitelnost dokumentů bez ohledu na zastarávání nebo změny technologií.	SIC, OIKT
Technický plán	Specifikuje požadavky na hardware, software a síťové systémy.	SIC, OIKT
Plán zajištění kontinuity	Zabývá se povinnostmi a zajištěním ochrany dokumentů i po zániku repozitáře.	SIC, OIKT, management univerzity
Krizový plán	Reaguje na náhlé změny v prostředí repozitáře.	SIC, OIKT, management univerzity

Tabulka 5: Přehled strategických cílů dle nástroje PLATTER s přidávanými zainteresovanými klíčovými skupinami

7 Závěr

V úvodní kapitole byla rozebrána problematika vývoje institucionálního repozitáře od prvotních různorodých systémů až po dnešní plně standardizovaný systém. Pokud se v těchto počátcích řešila otázka jak repozitář vybudovat, v dnešní době bychom se mohli ptát: Jak zajistit uloženým datům trvanlivost pro další generace. Dochovaly se nám tištěné akademické publikace staré několik staletí a dnes se řeší problém, jak zajistit stejnou životnost i digitálním dílům, která jsou mnohem zranitelnější. U tištěných publikací na sebe berou odpovědnost za řešení tohoto problému především knihovny, které jsou podporovány zákonem, konkrétně tzv. povinným výtiskem. Čistě digitální publikace, tedy publikace, které vznikají pouze v digitální formě, takto zálohované nejsou. Proto je na tvůrcích digitálního obsahu, aby archivaci zajistili. Řešením tohoto problému na institucionální úrovni mohou být důvěryhodné institucionální repozitáře.

Cílem této práce bylo představit aktuální pohled na problematiku důvěryhodných institucionálních repozitářů a nabídnout návrh postupu implementace takového repozitáře na České zemědělské univerzitě. Pojem institucionálního repozitáře byl tedy zúžen na oblast univerzitního institucionálního repozitáře. Byly zmapovány relevantní standardy, nástroje a doporučení, které mohou důvěryhodnost zaručit, nebo alespoň pomoci s jejím dosažením. Vývoj těchto standardů započal v roce 2002 vydáním modelu OAIS, který stanovil jednotnou terminologii pro repozitáře a stanovil první doporučení pro jejich tvůrce. V evropském prostoru, který je relevantní pro tuto práci, byl vývoj standardizace završen Evropským rámcem pro audit a certifikaci digitálních repozitářů. Tento dokument stanovil tři typy certifikace pro důvěryhodné repozitáře – základní certifikaci, rozšířenou certifikaci a formální certifikaci. Při sbírání poznatků o těchto certifikacích bylo pozorováno, že základní certifikace je nejčastější certifikací pro univerzitní či oborové repozitáře. Rozšířená certifikace je důkladnější a samozřejmě také náročnější na zdroje. Tato certifikace byla mnohem méně využívána a sáhly po ní repozitáře na národní či mezinárodní úrovni.

Nedílnou součástí institucionálního repozitáře jsou jeho politiky, mandát od vedení instituce a napojení na fungující institucionální informační infrastrukturu. Z toho důvodu byly porovnány politiky na českých institucionálních, oborových a národních repozitářích.

Implementace politik se velice liší z instituce na instituci, ale lze zmínit několik trendů: (1) Institucionální repozitáře jsou většinou, ne-li vždy, provozovány univerzitní knihovnou, případně knihovnou výzkumné instituce. (2) Převládá snaha zpřístupnit uložené záznamy v rámci otevřeného přístupu. (3) Institucionální repozitáře jsou většinou napojeny na fungující

univerzitní infrastrukturu, univerzitní informační systém. (4) Institucionální repozitáře se většinou nezavazují k dlouhodobé archivaci a neabsolvují ani alespoň základní certifikaci. (5) Institucionální repozitáře mají obecně nedostatečně popsány politiky svého fungování.

V rámci praktické části práce byl proveden průzkum mezi akademickými pracovníky České zemědělské univerzity za účelem zjištění rozdílů v postoji k otevřenému přístupu a motivaci k ukládání vlastních prací do institucionálního repozitáře. Z důvodu nízké návratovosti dotazníku se nepodařilo statisticky prokázat korelaci mezi postojem k otevřenému přístupu a akademickou hodností, oborem či předchozí zkušeností s publikováním v režimu otevřeného přístupu. Mezi nejkladněji hodnocené motivace pro vlastní ukládání respondenti uvedli možnost zobrazení statistik užívání u vlastních nahraných prací, formální uznání univerzitou či přínos vlastního ukládání do habilitačního a jmenovacího řízení či doktorského studia.

Na základě těchto zjištění a aplikace nástroje PLATTER byl poté vypracován návrh postupu pro tvorbu institucionálního repozitáře na České zemědělské univerzitě. Byly stanoveny klíčové skupiny, které je nutné do procesu plánování zapojit, a které mohou podobu výsledných plánů nějakým způsobem ovlivnit. Dále byly stanoveny základní parametry repozitáře v oblastech účelu a funkce, velikosti, provozu a technického řešení repozitáře. Nakonec byly vypsány plány strategických cílů s uvedenými zainteresovanými klíčovými skupinami, které by se na tvorbě těchto plánů měly podílet.

Absence institucionálního repozitáře na České zemědělské univerzitě by se v době, kdy každá větší univerzita v Česku již nějaký repozitář má, mohla jevit jako negativum. Nicméně na druhou stranu může Česká zemědělská univerzita z této pozice těžit a jít tzv. vyšlapanou cestou. Může se poučit z nedokonalostí a přešlapů ostatních a nakonec možná i vytvořit první doopravdy oficiálně důvěryhodný institucionální repozitář. Takový v Česku totiž zatím chybí.

8 Seznam použité literatury

- CARIFIO, James a Rocco PERLA, 2008. Resolving the 50-year debate around using and misusing Likert scales. In: *Medical Education* [online]. s. 1150–1152 [vid. 2017-05-04]. ISSN 03080110. Dostupné z: doi:10.1111/j.1365-2923.2008.03172.x
- CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. AUDIT AND CERTIFICATION OF TRUSTWORTHY DIGITAL REPOSITORIES [online]. 2011. [vid. 2016-04-08]. Dostupné z: <http://wiki.digitalrepositoryauditandcertification.org/pub/Main/WebHome/652x0r1candidate-update-typo-corrected.doc>
- ČESKÁ REPUBLIKA, 1998. Zákon č. 111 ze dne 22. dubna 1998 o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách) [online]. 1998. B.m.: Sbírka zákonů České republiky. 1998, částka 39, s. 5388-5419. ISSN 1211-1244. Dostupné z: <https://portal.gov.cz/app/zakony/zakonInfo.jsp?idBiblio=46613&nr=111~2F1998&rpp=15#local-content>
- ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA. Výroční zpráva o činnosti za rok 2016 [online]. 2017. Dostupné z: <https://www.czu.cz/dl/53958?lang=cs>
- ISO 14721:2012. *Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model* [online]. Geneva: International Organisation for Standardisation, 2012 [vid. 2017-06-29]. Dostupné z: <https://www.iso.org/standard/57284.html>
- ISO 16363:2012. *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories* [online]. Geneva: International Organisation for Standardisation, 2012 [vid. 2015-01-08]. Dostupné z: http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510
- KIM, Jihyun. Faculty self-archiving: Motivations and barriers. *Journal of the American Society for Information Science and Technology* [online]. 2010. **61**(9), 1909–1922 [vid. 2015-01-20]. Dostupné z: doi:10.1002/asi.21336
- KIM, Jihyun. Motivations of Faculty Self-archiving in Institutional Repositories. *The Journal of Academic Librarianship* [online]. 2011. **37**(3), 246–254. Dostupné z: doi:10.1016/j.acalib.2011.02.017
- MINISTERSTVO ŠKOLSTVÍ MLÁDEŽE A TĚLOVÝCHOVY. Výroční zpráva o stavu a rozvoji vzdělávání v České republice v roce 2015 [online]. Praha: Úřad vlády České republiky. 2015. 122 s. ISBN 978-80-87601-37-2. Dostupné z: <http://www.msmt.cz/file/36121/>
- NESTOR CERTIFICATION WORKING GROUP. *Explanatory notes on the nestor Seal for Trustworthy Digital Archives* [online]. 2013. Dostupné z: <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2013100901>
- PEJŠOVÁ, Petra. *Metodika zpracování, dlouhodobého uchování a zpřístupnění šedé literatury v ČR na příkladu Národního úložiště šedé literatury* [online]. Praha: Národní technická knihovna. 2011. 63 s. [vid. 2015-01-08]. Dostupné z: <http://www.nusl.cz/ntk/nusl-111907>
- ROSENTHAL, Colin., Asger. BLEKINGE-RASMUSSEN a Jan HUTAŘ. *Průvodce plánem důvěryhodného digitálního repozitáře (PLATTER)* [online]. Praha: Národní knihovna České republiky. 2009. 51 s. [vid. 2016-09-05]. ISBN 9788070505694. Dostupné z: <http://www.ndk.cz/platter-cz/Platter.pdf>

- ROSENTHAL, David S H, Thomas ROBERTSON, Tom LIPKIS, Vicky REICH a Seth MORABITO. Requirements for Digital Preservation Systems. *D-Lib Magazine* [online]. 2005. **11**(11) [vid. 2016-07-12]. Dostupné z: doi:10.1045/november2005-rosenthal
- RÖSSLEROVÁ, Klára. *Historie, současnost a budoucnost výměnných formátů bibliografických dat*. Praha. 2016. 190 s. Disertace. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví.
- SCHÖPFEL, Joachim, Stéphane CHAUDIRON, Bernard JACQUEMIN, Hélène PROST, Marta SEVERO a Florence THIAULT. Open access to research data in electronic theses and dissertations: an overview. *Library Hi Tech* [online]. 2014. **32**(4), 612–627. Dostupné z: doi:10.1108/LHT-06-2014-0058
- SMITH, MacKenzie, Mary BARTON, Margret BRANSCHOFKY, Greg MCCLELLAN, Julie Harford WALKER, Mick BASS, Dave STUVE a Robert TANSLEY. DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine* [online]. 2003. **9**(1). Dostupné z: doi:10.1045/january2003-smith
- THE CENTER FOR RESEARCH LIBRARIES a Inc. ONLINE COMPUTER LIBRARY CENTER. *Trustworthy Repositories Audit & Certification: Criteria and Checklist* [online]. 2007. Dostupné z: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
- THE CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. *Reference Model for an Open Archival Information System (OAIS)* [online]. 2012. Dostupné z: doi:<http://public.ccsds.org/publications/archive/650x0m2.pdf>
- THE DIGITAL CURATION CENTER a DIGITALPRESERVATIONEUROPE. *Digital Repository Audit Method Based on Risk Assessment* [online]. B.m.: Digital Curation Centre Partners and DigitalPreservationEurope Partners. 2007. 221 s. [vid. 2017-05-14]. ISBN 978-1-906242-00-8. Dostupné z: <http://www.repositoryaudit.eu/download>
- UNIVERSITY OF SOUTHAMPTON. *EPrints version 3 Repository Walkthrough* [online]. 2007. Dostupné z: http://www.eprints.org/software/v3/EPrintsv3Presentation_small.pdf
- VAN DE SOMPEL, Herbert a Carl LAGOZE. The Santa Fe Convention of the Open Archives Initiative. *D-Lib Magazine* [online]. 2000. **6**(2) [vid. 2017-02-19]. Dostupné z: <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>
- VASKA, Marcus, Joachim SCHÖPFEL, Iveta FÜRSTOVÁ, Radim POLČÁK, Jan MACH, Bohdana FRANTÍKOVÁ, Petr KARLACH a Jindřich DOLANSKÝ. *Repozitáře šedé literatury* [online]. Praha: VerBuM. 2010. 152 s. [vid. 2017-07-24]. ISBN 978-80-904273-5-8. Dostupné z: <https://nysl.techlib.cz/images/Book.pdf>
- WORKING GROUP ON TRUSTED REPOSITORIES CERTIFICATION. *2nd ed.: Catalogue of Criteria for Trusted Digital Repositories* [online]. 2009. Dostupné z: http://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf
- SPARC Institutional Repository Checklist and Resource Guide [online] CROW, Raym 2002. Dostupné z: www.arl.org/sparc/bm~doc/ir_guide__checklist_v1.pdf
- The Open Archives Initiative Protocol for Metadata Harvesting - v.2.0. *Open Archives Initiative* [online] LAGOZE, Carl, Herbert VAN DE SOMPEL, Michael NELSON a Simeon WARNER 2002 [vid. 2017-02-23]. Dostupné z: <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Institucionální repozitář. *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online] CUBR, Ladislav a Jaroslava HAVLOVÁ 2003 [vid. 2017-01-31]. Dostupné z: http://aleph.nkp.cz/F/?func=direct&doc_number=000014622&local_base=KTD

About the Registry. *Registry of Open Access Repositories* [online] UNIVERSITY OF SOUTHAMPTON 2004 [vid. 2017-02-09]. Dostupné z: <http://roar.eprints.org/information.html>

Důvěryhodný digitální repozitář. *Česká terminologická databáze knihovnictví a informační vědy* [online] CUBR, Ladislav a Jaroslava HAVLOVÁ 2006 [vid. 2017-02-27]. Dostupné z: http://aleph.nkp.cz/F/?func=direct&doc_number=000014621&local_base=KTD

Linked Data - Design Issues. *World Wide Web Consortium* [online] BERNERS-LEE, Tim 2010 [vid. 2017-03-30]. Dostupné z: <https://www.w3.org/DesignIssues/LinkedData>

Memorandum of Understanding. *Trusted Digital Repository* [online] CCSDS, DSA a NESTOR 2010. Dostupné z: [http://www.trusteddigitalrepository.eu/Memorandum of Understanding.html](http://www.trusteddigitalrepository.eu/Memorandum%20of%20Understanding.html)

History. *EPrints Documentation* [online] UNIVERSITY OF SOUTHAMPTON 2011 [vid. 2017-01-01]. Dostupné z: <http://wiki.eprints.org/w/History>

About OpenDOAR. *The Directory of Open Access Repositories* [online] UNIVERSITY OF NOTTINGHAM 2014 [vid. 2017-02-09]. Dostupné z: <http://www.opendoar.org/about.html>

JAK ZVEŘEJNIT ČLÁNEK V DIGITÁLNÍ KNIHOVNĚ VUT [online] ÚSTŘEDNÍ KNIHOVNA VUT V BRNĚ 2014 [vid. 2017-07-01]. Dostupné z: <https://www.vutbr.cz/uk/digitalni-knihovna/jak-zverejnit>

Produkty - ARL. *Cosmotron* [online] COSMOTRON BOHEMIA S.R.O. 2014 [vid. 2017-01-01]. Dostupné z: <http://www.cosmotron.cz/produkty/arl>

About. *Invenio Digital Library Framework Documentation* [online] CERN 2015a [vid. 2017-01-01]. Dostupné z: <http://invenio.readthedocs.io/en/latest/introduction/about.html>

O digitální knihovně FF MU. *Digitální knihovna Filozofické fakulty Masarykovy univerzity* [online] FILOZOFICKÁ FAKULTA MU 2015 [vid. 2017-07-02]. Dostupné z: <https://digilib.phil.muni.cz/pages/about>

Objectives of the DRAMBORA Process. *DRAMBORA Interactive* [online] THE DIGITAL CURATION CENTER a DIGITALPRESERVATIONEUROPE 2015 [vid. 2017-05-14]. Dostupné z: <http://www.repositoryaudit.eu/objectives/>

Podrobnosti o našem systému. *Informační systém Masarykovy univerzity* [online] FAKULTA INFORMATIKY MU 2015 [vid. 2017-01-01]. Dostupné z: https://is.muni.cz/nas_system/

Triaging. *Invenio Digital Library Framework Documentation* [online] CERN 2015b [vid. 2017-01-01]. Dostupné z: <http://invenio.readthedocs.io/en/latest/processes/triaging.html>

Governance. *DSpace* [online] DURASPACE ORGANIZATION 2016a [vid. 2017-01-01]. Dostupné z: <http://www.dspace.org/governance>

Návod na zadávání publikačních výsledků dle open access v systému OBD [online] UNIVERZITNÍ KNIHOVNA ZČU 2016 [vid. 2017-07-12]. Dostupné z: http://www.knihovna.zcu.cz/export/sites/knihovna/publikace/dokumenty-publikacni-cinnost/OA_a_evidence_OBD.pptx

Publikační činnost VŠB-TUO ve Web of Science [online] VŠB-TUO 2016 [vid. 2017-07-12].
Dostupné z: <https://dspace.vsb.cz/handle/10084/56138>

Technical Specifications. *DSpace* [online] DURASPACE ORGANIZATION 2016b [vid. 2017-01-01]. Dostupné z: <http://www.dspace.org/techspecs>

About LINDAT/CLARIN. *Centre for Language Research Infrastructure in the Czech Republic* [online] UFAL MFF UK 2017 [vid. 2017-06-16]. Dostupné z: <https://lindat.mff.cuni.cz/en/about-lindat-clarin>

arXiv Primer. *arXiv.org help* [online] CORNELL UNIVERSITY LIBRARY 2017a [vid. 2017-02-03]. Dostupné z: <https://arxiv.org/help/primer>

Assessment. *Data Seal of Approval* [online] DATA ARCHIVING AND NETWORKED SERVICES 2017 [vid. 2017-05-24]. Dostupné z: <https://www.datasealofapproval.org/en/assessment/>

Autoarchivace (zelená cesta OA) [online] UNIVERZITNÍ KNIHOVNA UNIVERZITY PARDUBICE 2017 [vid. 2017-07-04]. Dostupné z: <http://www.upce.cz/knihovna/vav/oa/green-oa.html>

Digitální knihovna ČVUT (repozitář) [online] ÚSTŘEDNÍ KNIHOVNA ČVUT 2017 [vid. 2017-07-01]. Dostupné z: <http://knihovna.cvut.cz/cs/vyhledavani/hledat-v/digitalni-knihovna-repozitar>

General Information About arXiv. *arXiv.org help* [online] CORNELL UNIVERSITY LIBRARY 2017b [vid. 2017-03-02]. Dostupné z: <https://arxiv.org/help/general>

Institucionální digitální repozitář [online] NÁRODNÍ TECHNICKÁ KNIHOVNA 2017a [vid. 2017-07-05]. Dostupné z: <https://www.techlib.cz/cs/2984-institucionalni-digitalni-repozitar>

Institucionální digitální repozitář NTK. *Národní technická knihovna* [online] NÁRODNÍ TECHNICKÁ KNIHOVNA 2017b [vid. 2017-01-01]. Dostupné z: <http://repozitar.techlib.cz/?ln=cs>

Pravidla repozitáře [online] KNIHOVNA AKADEMIE VĚD ČR 2017a [vid. 2017-07-12]. Dostupné z: <https://asep.lib.cas.cz/arl-cav/cs/pravidla-repozitare/>

Repozitář AV ČR [online] KNIHOVNA AKADEMIE VĚD ČR 2017b [vid. 2017-07-12]. Dostupné z: <https://www.lib.cas.cz/asep/uvod-do-asep/repozitar-avcr/>

Repozitáře [online] KNIHOVNA UK V PRAZE 2017 [vid. 2017-06-20]. Dostupné z: <https://knihovna.cuni.cz/rozcestnik/repozitare/>

SRW/U. *OCLC Research* [online] OCLC 2017 [vid. 2017-01-01]. Dostupné z: <http://www.oclc.org/research/themes/data-science/srw.html>

Technické řešení NUŠL [online] NUŠL 2017 [vid. 2017-07-05]. Dostupné z: <https://nusl.techlib.cz/nusl/technicke-reseni/>

Univerzitní repozitář [online] FI MU 2017 [vid. 2017-07-02]. Dostupné z: <https://is.muni.cz/repozitar/>

Zpřístupňování vysokoškolských kvalifikačních prací (VŠKP) VŠB-TUO [online] ÚSTŘEDNÍ KNIHOVNA VŠB-TUO 2017 [vid. 2017-07-12]. Dostupné z: <http://knihovna.vsb.cz/sluzby/sluzbystudovna.htm#vskp>

Seznam ilustrací a tabulek

Ilustrace

Ilustrace 1: Schéma protokolu OAI-PMH	14
Ilustrace 2: Ukázka popisu dokumentu za pomoci schématu Dublin Core v hlavičce dokumentu.	16
Ilustrace 3: Přehled softwaru použitého pro institucionální repozitáře v České republice podle OpenDOAR	19
Ilustrace 4: Graf softwaru použitého pro repozitáře v České republice podle registru ROAR	20
Ilustrace 5: Diagram faktorů a jejich předpokládaného pozitivního či negativního vlivu na chování self-archivingu.	56
Ilustrace 6: Výsledný model hlavních faktorů majících vliv na self-archiving	57

Tabulky

Tabulka 1: Seznam kritérií certifikace Nestor Seal for Trustworthy Digital Archives a jejich minimální bodové ohodnocení	39
Tabulka 2: Shrnutí politik institucionálních repozitářů v České republice	43
Tabulka 3: Schéma výpočtu pozorovaných četností a očekávaných četností pro jednotlivé obory	59
Tabulka 4: Přehled odpovědí na otázky týkající se motivace ukládání prací do repozitáře....	63
Tabulka 5: Přehled strategických cílů dle nástroje PLATTER s přidáním zainteresovanými klíčovými skupinami	69

Příloha 1 – Otázky dotazníku pro vědecké pracovníky na ČZU

Výzkum vztahu vědců k otevřenému přístupu a vlastní archivaci na České zemědělské univerzitě

Vážení kolegové a kolegyně,
tento dotazník má za úkol zmapovat vztah vědců na České zemědělské univerzitě k otevřenému přístupu a vlastní archivaci vědeckých dokumentů. O vytvoření dotazníku a jeho zpracování se stará Studijní a informační centrum České zemědělské univerzity.
Otázky dotazníku jsou sestaveny tak, aby byla zaručena anonymita respondentů.

***Povinné pole**

1. Prosím vyberte váš věk *

Označte jen jednu elipsu.

- ☐ 20-30
☐ 31-40
☐ 41-55
☐ 55+

2. Prosím uveďte vaši akademickou hodnost. *

Označte jen jednu elipsu.

- ☐ Magistr / Inženýr
☐ Doktor
☐ Docent
☐ Profesor
☐ Jiné: _____

3. Prosím vyberte váš obor *

Označte jen jednu elipsu.

- ☐ Jiný (specifikujte v následující otázce)
- ☐ Architektura a urbanismus
- ☐ Biologie a biochemie
- ☐ Cestování, turismus a volný čas
- ☐ Ekonomie
- ☐ Elektronika a automatizace
- ☐ Elektrotechnika a energetika
- ☐ Fyzika
- ☐ Hornictví a těžba
- ☐ Chemické výroby
- ☐ Chemie
- ☐ Lesnictví
- ☐ Management a správa
- ☐ Marketing a propagace
- ☐ Matematika
- ☐ Motorová vozidla, lodě a letadla
- ☐ Ochrana životního prostředí
- ☐ Pedagogika
- ☐ Peněžnictví, bankovnictví, pojišťovnictví
- ☐ Počítačové vědy
- ☐ Politické vědy a občanská nauka
- ☐ Potravinářství
- ☐ Pracovní prostředí
- ☐ Právo
- ☐ Přepavní služby a spoje
- ☐ Příprava učitelů a učení
- ☐ Přírodní prostředí a přirozené formy života
- ☐ Psychologie
- ☐ Rostlinná a živočišná výroba
- ☐ Rybářství
- ☐ Sociologie a kulturologie
- ☐ Sporty
- ☐ Statistika
- ☐ Stavebnictví a inženýrské stavitelství
- ☐ Strojírenství, kovovýroba a metalurgie
- ☐ Technologie ochrany životního prostředí
- ☐ Účetnictví a daně
- ☐ Užití počítačů
- ☐ Vědy o Zemi
- ☐ Vědy o životním prostředí
- ☐ Velkoobchod a maloobchod

- ☐ Veterinářství
- ☐ Výroba a zpracování materiálu (dřevo, papír, plasty, sklo)
- ☐ Výroba textilu, oděvu a obuvi, zpracování kuže
- ☐ Zahradnictví
- ☐ Žurnalistika a zpravodajství
- ☐ Jiný (specifikujte v následující otázce)

4. Jiný obor

5. Vyberte z následujících činností tu, která nejvíce odpovídá úrovni Vašich technických dovedností s počítačem. *

Označte jen jednu elipsu.

- ☐ Umím vypnout a zapnout počítač.
- ☐ Dovedu si nainstalovat program na počítač.
- ☐ Umím si vytvořit PDF z Word dokumentu.
- ☐ Dokážu nahrát článek do uživatelského rozhraní časopisu.
- ☐ Umím si nainstalovat operační systém.
- ☐ Jiné: _____

6. Ročně publikuji přibližně *

Označte jen jednu elipsu.

- ☐ Žádnou publikaci
- ☐ 1-2 publikace
- ☐ 3-5 publikací
- ☐ 6-9 publikací
- ☐ 10 a více publikací

7. Už jsem někdy zpřístupnil plný text své práce veřejně skrze nějakou jinou instituci než vydavatele časopisu? *

Zaškrtněte všechny platné možnosti.

- ☐ Ano, v oborovém repozitáři (RePEc, arXiv.org, CogPrints, CiteSeerX, apod.)
- ☐ Ano, na osobním nebo fakultním webu
- ☐ Ano, na sociální síti jako ResearchGate či Mendeley
- ☐ Ano, v institucionálním repozitáři univerzity.
- ☐ Ne
- ☐ Jiné: _____

8. Pokud v oborovém repozitáři, ve kterém?

Zaškrtněte všechny platné možnosti.

- ☐ RePEc
☐ arXiv.org
☐ CogPrints
☐ CiteSeerX
☐ Jiné: _____

Váš vztah k otevřenému přístupu a vlastní archivaci vaší vědecké práce.

Na otázky v této sekci můžete odpovídat na škále Zcela nesouhlasím (1) - Spíše nesouhlasím (2) - Nemám názor (3) - Spíše souhlasím (4) - Zcela souhlasím (5)

Institucionální repozitář

je informační systém pro ukládání vědeckého výstupu zaměstnanců a studentů instituce. Může se jednat o rukopisy odborných textů, kvalifikační práce studentů apod.

Publikování v režimu otevřeného přístupu

je způsob zveřejňování vlastní práce tak, aby k ní měl přístup kdokoli, tj. bez jakýchkoli poplatků a nutnosti se registrovat. Mohou se zveřejňovat jak rukopisy prací, tak i hotové články.

9. 1. Myslím si, že princip otevřeného přístupu, kdy má k vědeckým materiálům přístup kdokoli, je správný. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

10. 2. Volné zpřístupňování mých vědeckých materiálů na internetu pomůže ostatním vědcům stavět na mých vědeckých zjištěních. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

11. 3. Volné zpřístupňování mých vědeckých materiálů na internetu umožní ostatním vědcům získat přístup k materiálům, ke kterým by se jinak nedostali. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

12. 4. V mém oboru je běžné, že vědci vystavují nerecenzované či recenzované verze svých článků veřejně na internetu v režimu volného přístupu. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

13. 5. Na mé katedře je běžné, že vytváříme a sdílíme mezi sebou návrhy článků (working paper) nebo technické zprávy. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

14. 6. Abych mohl zveřejnit svůj článek na internetu v režimu otevřeného přístupu, musím požádat o svolení vydavatele časopisu, kde jsem článek publikoval. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

15. 7. Pokud zpřístupním svůj článek na internetu v režimu otevřeného přístupu bez vyžádání povolení vydavatele mého článku, porušuji tím autorské právo (příp. copyright). *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

16. 8. K publikování mého článku v režimu otevřeného přístupu na internetu potřebuji souhlas všech spoluautorů. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

17. 9. Nemohu publikovat svůj článek v časopise, pokud jsem ho předtím zveřejnil v režimu otevřeného přístupu na internetu. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

18. 10. Na mé fakultě je výzkumná práce zveřejněná v režimu otevřeného přístupu akceptovaná jako alternativa k publikaci v časopise pro účely doktorského studia a habilitačního a jmenovacího řízení. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

19. 11. Zveřejnění mého článku v režimu otevřeného přístupu negativně ovlivní mé šance v doktorském studiu a habilitačním či jmenovacím řízení. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

20. 12. Zveřejnění mého článku v režimu otevřeného přístupu negativně ovlivní mé šance při získávání výzkumných grantů. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

21. 13. Ke zveřejnění mých článků v režimu otevřeného přístupu je potřeba vynaložit značné úsilí a čas. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

22. 14. Je obtížné naučit se, jak zadávat požadovaná data o článku (název, autoři, klíčová slova,...) při nahrávání do repozitáře či na internet. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

23. 15. Zveřejňování mých článků v režimu otevřeného přístupu mi ubere čas, který bych jinak využil k výzkumu a psaní. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

Co by vás motivovalo k nahrání vaší práce do institucionálního repozitáře?

24. **16. Kdyby dokázal spolehlivě uchovat uložené soubory po mnoho let. (15 let a více) ***

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

25. **17. Kdyby bylo vidět, kolikrát si někdo moji práci zobrazil či stáhnul. ***

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

26. **18. Kdyby nahrávání probíhalo automaticky nebo by se o to za mě staral někdo jiný. ***

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

27. **19. Že by vydavatel nebyl jediný, kdo má kontrolu nad zpřístupněním mé práce. ***

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

28. **20. Kdyby univerzita nějak formálně ocenila můj přínos do repozitáře. ***

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

29. **21. Možnost ukázat můj výzkum, ještě než by byl dokončen, pro případnou kritiku od kolegů. ***

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

30. **22. Kdyby se přispívání do IR počítalo do habilitačního a jmenovacího řízení či doktorského studia. ***

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

31. 23. Kdybych dostal/a finanční odměnu. *

Označte jen jednu elipsu.

	1	2	3	4	5	
Zcela nesouhlasím	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Zcela souhlasím

Komentář k dotazníku

32. Komentář k dotazníku

Napadlo Vás něco během vyplňování dotazníku? Vzkazte to autorovi.

Používá technologii



Google Forms

Příloha 2 – Výpočet chí-kvadrátu pro vyloučení H_0

Vztah mezi oborem a nadprůměrností či podprůměrností OA skóre									
P... Pozorované četnosti									
O... Očekávané četnosti podle H0									
		Nadprůměrný OA							
Obor	P NEPRAVDA	O Nepravda	P PRAVDA	O Pravda		Celkový součet			
Přírodní vědy, matematika a informatika	7	7		6	6	13			
Společenské vědy, obchod a právo	1	0,538461538		0	0,461538462	1			
Technika, výroba a stavebnictví	1	1,615384615		2	1,384615385	3			
Zemědělství a veterinářství	5	4,846153846		4	4,153846154	9			
Celkový součet	14	14		12	12	26			
	0			0					
	0,030431107			0,030431107					
	0,054099746			0,054099746					
	0,003381234			0,003381234					
				Chí-kvadrát	0,175824176				
Kritická hodnota pro hladinu významnosti 0,05 pro 4 řádky a dva sloupce (3 stupně volnosti) dle tabulek kritických hodnot pro chí-kvadrát									
						7,815			